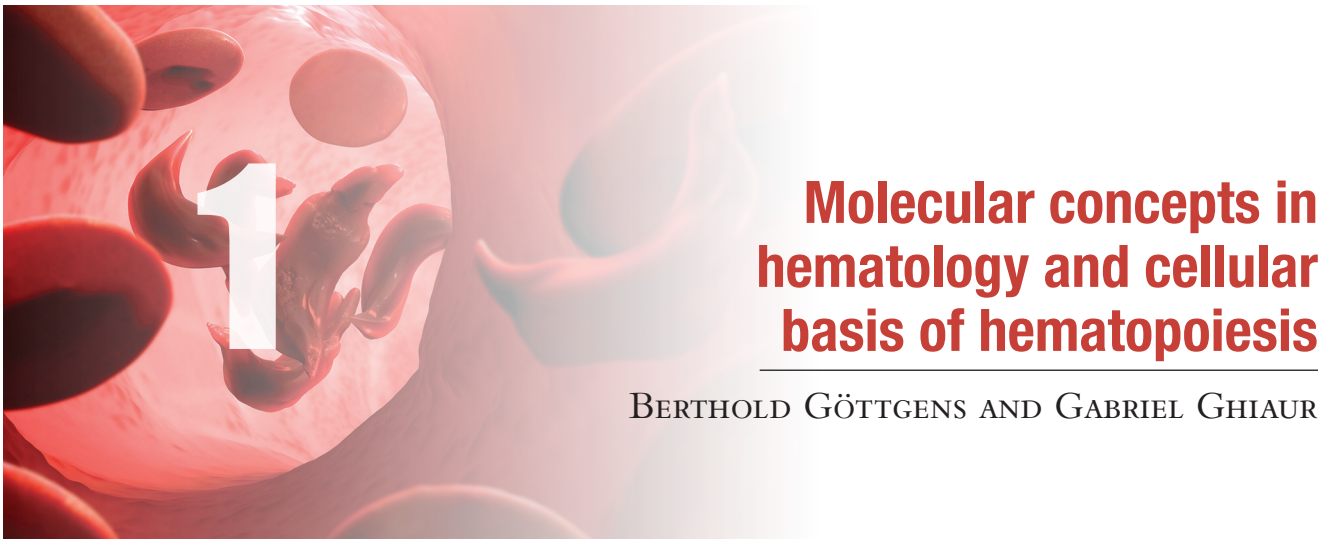


Molecular concepts in hematology and cellular basis of hematopoiesis

Berthold Göttgens and Gabriel Ghiaur

Overview 1
Basic concepts of molecular biology 1
Analytic techniques 6
Clinical applications of DNA technology in hematology 14
Cellular basis of hematopoiesis 18
Glossary 29
Bibliography 32

The online version of this chapter contains an educational multimedia component on normal hematopoiesis.

Conflict-of-interest disclosure: Berthold Göttgens declares no competing financial interest. Gabriel Ghiaur declares no competing financial interest.

Off-label drug use: Berthold Göttgens: not applicable. Gabriel Ghiaur: not applicable.

Overview

The last several decades have seen major advances in experimental models and molecular techniques. These advances not only resulted in a deeper understanding of blood physiology but also provided diagnostic and therapeutic tools for blood-related diseases. This Chapter intends to familiarize hematologists with the molecular biology techniques that are finding their way in clinical practice and with the concepts of hematopoiesis that have direct implications for normal and malignant hematology.

Basic concepts of molecular biology

Central dogma

The gene

The central dogma of biology describes the flow of genetic information in cells from DNA to mRNA to proteins (Figure 1-1). The functional genetic unit responsible to produce a given protein is termed a **gene**. The gene has 4 major parts: (1) **exons**, which contain the **coding sequence** that will ultimately dictate the amino acids found in the proteins; (2) intervening **noncoding sequences**, **introns**; and (3) 2 **flanking regions**, 5′ and 3′, containing regulatory elements that control the level as well as spatial and temporal aspects of gene expression.

Genes are arrayed in a linear fashion along **chromosomes**, which are long DNA structures complexed with protein. Within chromosomes, DNA is bound in **chromatin**, a complex of DNA with histone and nonhistone proteins that "shield" the DNA from the proteins that activate gene expression. The ends of chromosomes are capped by **telomeres**, repetitive DNA sequences and their associated proteins. The DNA replication machinery cannot effectively replicate the very ends of chromosomes, and thus shortening occurs with each cell division, ultimately leading to chromosomal instability and cellular senescence. Telomeres are therefore critical to protecting chromosome ends from degradation and fusion. Mutations in genes encoding components of the telomere complex are associated with the bone marrow failure syndrome dyskeratosis congenita. Furthermore, inappropriate activation of the enzymes that maintain

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

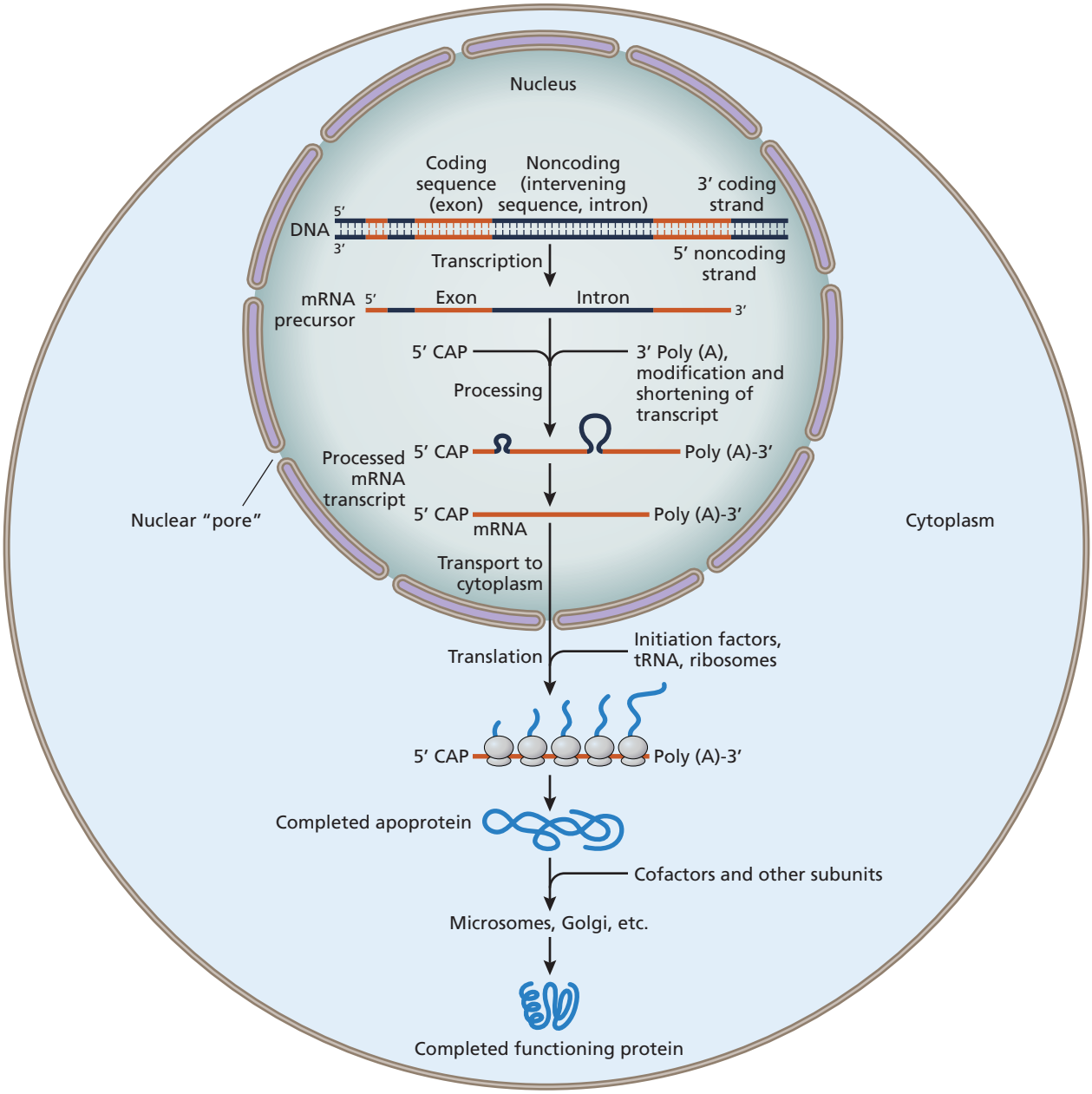


Figure 1-1 Flow of genetic information from DNA to RNA to protein. DNA is shown as a double-stranded array of alternating exons (red) and introns (blue). Transcription, posttranscriptional processing by splicing, polyadenylation, and capping are described in the text. The mature transcript passes from the nucleus to the cytoplasm, where it is translated and further modified to form a mature protein. Reproduced from Hoffman R et al, eds., Hematology: Basic Principles and Practice, 6th ed. (Saunders Elsevier; 2013:5), with permission.

telomeres ("telomerase") is associated with numerous cancers, including hematologic malignancies.

The RNA

The first step in the expression of protein from a gene is the synthesis of a **pre-messenger RNA** (pre-mRNA). The **transcription** of pre-mRNA is directed by **RNA polymerase II**, which, in conjunction with other proteins, generates an RNA copy of the DNA sense strand. The introns are then removed by a complex process called mRNA **splicing**. This process involves the recognition of specific sequences near either end of the intron that allow its excision in a precise manner that maintains the exon sequence. The mRNA may then undergo modifications at the 5′ and 3′ ends (**capping** and **polyadenylation**, respectively). Splicing of mRNA is a critical step

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

in gene expression with important implications for understanding hematologic disease. Splicing is controlled by the spliceosome, a large complex of proteins (100 to 300) and five small nuclear ribonuclear proteins (snRNPs). mRNA splicing is an important mechanism for generating diversity in the proteins produced by a single gene. Some genes exhibit **alternative splicing**, a process by which certain exons are included in or excluded from the mature mRNA, depending on which splice sequences are used in the excision process. This permits functional diversity of the products of the same gene and is one of several determinants of tissue specificity of cellular proteins.

Mutations in the sequences of either introns or exons can derange the splicing process by either creating or destroying a splice site so that the intron sequence is not removed or the exon sequence eliminated. If abnormal splicing results in a premature stop codon (**nonsense mutation**), then a surveillance pathway known as **nonsense-mediated decay** leads to degradation of the abnormal mRNA. Mutations affecting the spliceosome machinery are found in a variety of hematopoietic malignancies. For example, splicing factor mutations (SF3B1, U2AF1, and SRSF2) are found in approximately 10% of patients with chronic lymphocytic leukemia and 50% of patients with myelodysplastic syndrome (MDS), as well as related myeloid malignancies.

The protein

The mature mRNA is transported from the nucleus to the cytoplasm, where it undergoes **translation** into protein. The mRNA is "read" in a linear fashion by **ribosomes** that move along the mRNA and insert the appropriate amino acids, carried by **transfer RNAs (tRNAs)**, into the nascent protein. The amino acids are encoded by three base triplets called **codons**, the **genetic code**.

Single-base-pair alterations in the coding sequence of genes (point mutations) may have a range of effects on the resultant protein. Some single-base-pair changes may not alter the amino acid sequence or they may change the amino acid sequence in a manner that has no effect on the overall function of the protein; these are predicted to be phenotypically silent mutations. In other cases, mutations may lead to a loss or gain of protein function, or may result in the acquisition of a new function (**missense mutation**). Sickle cell disease is an example of a single-base-pair change, resulting in an amino acid alteration that critically changes the chemical characteristics of the globin molecule. Other mutations may change a codon to a **termination codon**, resulting in premature termination of the protein (**nonsense mutation**). Finally, single or multiple base-pair insertions or deletions can disrupt the reading frame of genes. These **frameshift mutations** render the gene incapable of encoding normal protein. These latter two abnormalities account for some β-thalassemias and for polycythemia due to a gain of function in the erythropoietin receptor. Clinically important mutations also may occur in the noncoding region of genes, such as in the regulatory elements upstream of the initiation codon or within intronic splicing sites.

Control of gene expression

Regulatory elements

With the exception of lymphocytes (which undergo unique changes in the DNA encoding immunoglobulin or the T-cell receptor) and germ cells (which contain only half of the DNA of somatic cells), each nucleated cell in an individual has the same diploid DNA content. Consequently, biologic processes are critically dependent on **gene regulation**, the control of gene expression such that proteins are produced only at the appropriate time within the appropriate cells. Gene regulation is the result of a complex interplay of specific sequences within a gene locus, chromatin, and regulatory proteins (transcription factors [TFs]) that interact with those sequences to increase or decrease the transcription from that gene.

DNA sequences that lie in proximity to and regulate the expression of genes, which encode protein, are termed ***cis*-acting regulatory elements**. Nearly all genes have a site for binding RNA polymerase II that is within the first 50 bases 5′ to the structural gene and is called the **promoter** region. Other sequences that regulate the level of transcription of the gene are located at less predictable distances from the structural gene. Such sequences may increase (**enhancers**) or decrease (**silencers**) expression. A special type of enhancer is the locus control region, which was first defined in the β-globin cluster of genes on chromosome 11. It is located approximately 50 kilobases (kb) upstream from the β-globin gene, controls all genes within the β-globin locus, and also has a strong tissue-specific activity (erythroid-specific).

Control of gene expression is exerted through the interaction of the *cis*-acting elements described previously with proteins that bind to those sequences. These nuclear DNA-binding proteins are termed ***trans*-acting factors** or **transcription factors**. Most of these proteins have a DNA-binding domain that can bind directly to regulatory sequences within the gene locus; many of them contain common motifs, such as **zinc fingers** or **leucine zippers**, which are shared by many transcription factors. In addition, they frequently have unique domains that allow them to interact with other transcription factors. Thus, a complex

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

pattern exists whereby different transcription factors may interact both with one another and with specific regions of DNA to increase or decrease transcription, and thus determine the unique tissue and stage-specific expression of the genes within a given cell.

Epigenetics

Epigenetic processes or modifications to the genome regulate gene expression without altering the underlying **nucleotide** sequence. These changes may be modulated by environmental factors and may be heritable. Two common forms of epigenetic changes are DNA methylation and histone modifications.

DNA methylation

DNA is modified by the addition of methyl groups to cytosine residues (resulting in 5-methylcytosine) through enzymes called DNA methyltransferases. **Methylation** normally occurs throughout the genome and is associated with alterations in gene expression and processes such as X-chromosome inactivation, imprinting, aging, and carcinogenesis. For instance, in females, DNA methylation results in X-chromosome inactivation. While either one of the two X chromosomes can be initially silenced, all daughter cells will continue to maintain the same silenced X chromosome. Using this principle in tumor tissue derived from females led to early demonstrations that neoplastic diseases are, for the most part, **clonal**.

Methylation of the 5′ regulatory sequences of the gene is generally a marker of an inactive gene. Changes in gene expression often can be correlated with the degree of methylation. In contrast to 5′ regulatory sequences where methylation is typically associated with gene silencing, methylation of intragenic cytosine residues often coincides with active transcription of the gene within which they lie.

Mendelian genetics is based on the principle that the phenotype is the same whether an **allele** is inherited from the mother or the father, but this does not always hold true. Some human genes are transcriptionally active on only one copy of a chromosome (such as the copy inherited from the father), whereas the other copy on the chromosome inherited from the mother is transcriptionally inactive. This mechanism of gene silencing is known as **imprinting**, and these transcriptionally silenced genes are said to be "imprinted." When genes are imprinted, they are usually heavily methylated in contrast to the nonimprinted copy of the allele, which is typically not methylated. A classic example of imprinting is the inheritance of Prader-Willi and Angelman syndromes.

As DNA methylation modulates gene activity, acquired changes in methylation patterns are thought to contribute to functional alterations in hematopoietic stem cells (HSCs) during aging. Hematopoietic stem cells display an age-related decline in function and a relative loss of lymphoid differentiation potential, and these changes are associated with site-specific changes in DNA methylation. Furthermore, aberrant methylation may contribute to cancer. For example, mutations in enzymes affecting DNA methylation, most notably the DNA methyltransferase DNMT3A, are common in acute myeloid leukemia (AML). Mutations in DNMT3A are found in approximately 20% of adult AML patients and are associated with poor outcomes. These mutations appear to be acquired early in the evolution of the disease, thus likely serving as initiators of the disease. Other enzymes affecting DNA methylation, including TET2 and IDH1/2, are also recurrently mutated in AML, further illustrating the role of aberrant methylation in the pathogenesis of this disease. Of note, oxidation of methylated cytosines (5-methylcytosine) by members of the ten-eleven translocation (TET) protein family results in the formation of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. While the role of these epigenetic marks is not clearly understood, their levels are often altered in cancers (including AML), suggesting that they may contribute to the pathogenesis of disease. Given the prominent role of abnormal methylation in hematologic malignancies, small molecule inhibitors of DNA methyltransferases (eg, 5-azacitidine, decitabine) are used in the treatment of disorders that are characterized by aberrant DNA methylation (eg, MDS, AML).

Histone modification

Histones are DNA packaging proteins that organize DNA into structural units called nucleosomes. Octamers of the core histones—H2A, H2B, H3, and H4—make up the nucleosome around which 147 base pairs (bp) of DNA is wrapped, and histone H1 binds the "linker" DNA between nucleosomes. Histones are subject to multiple modifications—including methylation, acetylation, ubiquitination, phosphorylation, and others. The particular combination of histone modifications at any given locus is thought to confer a "histone code," regulating processes such as gene expression, chromosome condensation, and DNA repair. Like methylation, histone modifications can regulate gene activity, and therefore disruptions of the normal pattern of these modifications can contribute to cancer and other diseases. For example, hypoacetylation of histones H3 and H4 is associated with silencing of the cell cycle regulator $p21^{WAF1}$, a gene whose expression is reduced in multiple tumor types. Aberrant expression levels of histone deacetylases, the enzymes that remove acetyl

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

groups from histone tails, are common in hematopoietic malignancies, and histone deacetylase inhibitors are being tested in a variety of these diseases. The first of these inhibitors, vorinostat, is used in the treatment of cutaneous T-cell lymphoma. In addition to altered expression levels of histone-modifying enzymes, several fusion genes associated with hematologic malignancies aberrantly recruit histone modifiers to target genes, resulting in altered transcription (eg, the association of MLL fusion proteins with the DOT1L enzyme that methylates histones and the histone acetyltransferase p300 with AML1-ETO). Targeting histone modifiers therapeutically therefore has potential utility for a variety of cancers.

Noncoding RNAs

It has been estimated that only approximately 1% to 2% of the genome encodes protein, but a much larger fraction is transcribed. This transcribed RNA that does not encode protein is referred to as noncoding RNA (ncRNA) and is grouped into an increasingly large number of different classes, including **microRNAs (miRs)**, small nucleolar RNAs (snoRNAs), **small interfering RNAs (siRNAs)**, Piwi-interacting RNAs (piRNAs), long noncoding RNAs (lncRNAs), and many others. Although each of these classes of ncRNAs differs in size, biogenesis pathway, and specific function, they share a common ability to recognize target nucleotide sequences through complementarity and regulate gene expression. The most well-described class of ncRNAs are microRNAs, whose biogenesis pathway is illustrated in Figure 1-2. Following transcription, a portion of this RNA (the pri-microRNA) forms hairpin loops that are cleaved by the enzymes **Drosha** and **Dicer** into short 21- to 23-bp double-stranded RNAs. These short double-stranded RNAs contain both sense strands and antisense strands that correspond to coding sequences in mRNAs. These mature miRs then are incorporated into a larger complex known as an **RNA-induced silencing complex (RISC)**. The miR is then unwound in a strand-specific manner, and the single-stranded RNA locates mRNA targets by Watson-Crick base pairing. Gene silencing results from cleavage of the target mRNA (if there is complementarity at the scissile site) or translational inhibition (if there is a mismatch at the scissile site). This gene-silencing pathway is known as RNA interference. As mediators of gene expression, miRs and other ncRNAs are expressed in a tissue-specific manner and play important regulatory roles in development and differentiation. Accordingly, dysregulation or mutations in ncRNAs are associated with various diseases, including cancer. In hematopoiesis, numerous miRs that influence

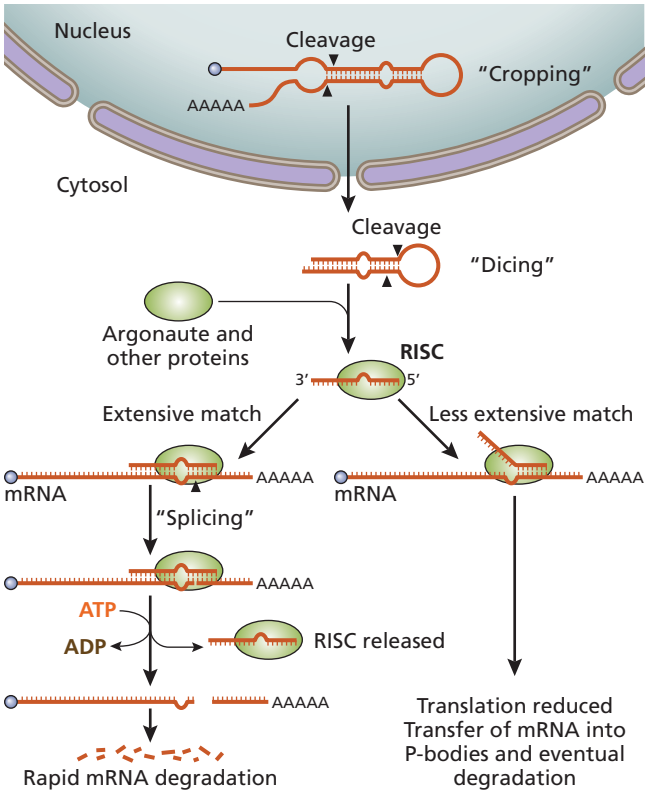


Figure 1-2 MicroRNA production. Production of microRNA begins with transcription of the microRNA gene to produce a stem-loop structure called a pri-microRNA. This molecule is processed by Drosha ("cropping") to produce the shorter pre-microRNA. The pre-microRNA is exported from the nucleus; the cytoplasmic Dicer enzyme cleaves the pre-microRNA ("dicing") to produce a double-stranded mature microRNA. The mature microRNA is transferred to RISC (RNA-induced silencing complex), where it is unwound by a helicase. Complementary base pairing between the microRNA and its target mRNA directs RISC to destroy the mRNA (if completely complementary) or halt translation (if a mismatch exists at the scissile site). Reproduced from Alberts B et al, Molecular Biology of the Cell, 5th ed. (Garland Science; 2007), with permission.

cell fate decisions have been identified. miR-223, for example, plays a central role in myeloid differentiation, and reduced expression of miR-223 is common in AML. Conversely, the expression of miR-125b, which is expressed in normal hematopoietic stem cells and whose targets include genes involved in regulating cell proliferation, differentiation, and survival, is increased in a variety of myeloid and lymphoid malignancies.

Molecular basis of neoplasia

Normal cellular growth and differentiation depends on the precise control of gene expression, and alterations in the quantity or timing of gene expression can affect the survival and function of a cell. When such alterations

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

occur in certain types of genes known as **oncogenes** or **tumor suppressor** genes, the cell may gain abnormal growth or survival properties, and accumulations of such mutations may lead to cancer.

Oncogenes

Oncogenes are genes that have the potential to cause cancer, and they arise from mutations in their normal counterparts termed **proto-oncogenes**. Proto-oncogenes generally code for proteins or ncRNAs that regulate such processes as proliferation and differentiation, and activating mutations or epigenetic modifications that increase the expression or enhance the function of these genes confers a growth or survival advantage on a cell. The first described oncogene, termed *SRC*, was discovered in the 1970s and is a member of a family of tyrosine kinases that regulate cell proliferation, motility, adhesion, survival, and differentiation. Activating mutations in the SRC family kinases are associated with the pathogenesis of multiple types of neoplasias, including cancers of the colon, breast, blood, head and neck, and others. Another classic example of an oncogene is the *BCR-ABL1* fusion gene found in chronic myelogenous leukemia (CML). This fusion results from a translocation between the *BCR* gene on chromosome 9 and the *ABL1* proto-oncogene on chromosome 22 and confers constitutive activation of *ABL1* and enhanced cell proliferation. Pharmacologic targeting of the activity of oncogenes, such as the use of the tyrosine kinase inhibitor imatinib to treat CML, can be an effective therapeutic approach.

Tumor suppressors

In contrast to oncogenes, tumor suppressors are genes that encode proteins or ncRNAs whose normal function is to inhibit tumor development through the promotion of such processes as apoptosis, DNA repair, cell cycle inhibition, cell adhesion, and others. Loss of the expression or function of these genes is associated with cancer, and generally both copies of the tumor suppressor gene must be altered to promote neoplasia. Thus, most tumor suppressors follow the "two-hit hypothesis" proposed by Alfred Knudson in his study of the retinoblastoma-associated tumor suppressor gene *RB1*. This gene encodes a protein that functions to regulate cell cycling and survival. Because both copies of the gene must be mutated for retinoblastoma to manifest, individuals that inherit a mutant allele (requiring just one more "hit" in the remaining normal allele for loss of gene function) generally develop disease earlier than those that must acquire "hits" in both alleles. Familial cancer syndromes often result from the inheritance of heterozygous mutations in tumor suppressor genes. For example, Li-Fraumeni syndrome results from inherited mutations in the cell cycle regulator *TP53* and is associated with the early onset of multiple tumor types, including osteosarcoma, breast cancer, leukemia, and others. When mutations occur in the remaining normal allele, termed "loss of heterozygosity," tumor growth is initiated. In some cases, loss of just one copy of a gene ("haploinsufficiency") has been shown to contribute to cancer development. For example, loss of one copy of the ribosomal gene RPS14 in patients with 5q- syndrome leads to aberrant ribosomal protein function and a block in erythroid differentiation.

Analytic techniques

Polymerase chain reaction

The polymerase chain reaction (PCR) is a powerful technique for amplifying small quantities of DNA of known sequence. Two **oligonucleotide** primers are required: one is **complementary** to a sequence on the 5′ strand of the DNA to be amplified and the other is complementary to the 3′ strand. The double-stranded DNA template is denatured into single strands at high temperature; the temperature then is lowered for the primers to be annealed to the DNA. The DNA then is extended with a temperature-stable DNA polymerase (such as *Taq* polymerase), resulting in two identical copies of the original DNA from each piece of template DNA. The products are denatured, and the process is repeated through multiple cycles, resulting in exponential amplification. The primary product of this reaction is the fragment of DNA bound by the two primers. Thus, small quantities of input DNA may be used to synthesize large quantities of a specific DNA sequence. This technique has superseded many blotting techniques (eg, the so-called Southern blot) for prenatal diagnosis and cancer diagnostics. Using multiple primer pairs in the same reaction, multiplex PCR can efficiently amplify several fragments simultaneously.

Reverse transcriptase PCR (RT-PCR) is a modification of the PCR technique that allows the detection and amplification of expressed RNA transcripts. **Complementary DNA (cDNA)** is generated from RNA using reverse transcriptase, an enzyme that mediates the conversion of RNA to DNA. The resultant cDNA is then subjected to routine PCR amplification. Because cDNA is generated from processed mRNA transcripts, no intronic sequences are obtained. RNA is much less stable than DNA; thus, amplification of mRNA from tissue or blood requires careful preservation of source tissue or blood samples.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

Quantitative PCR is another modification of the PCR technique. The most commonly used method is **real-time PCR**, in which a fluorogenic tag is incorporated into an oligonucleotide that will anneal to the internal sequence of the *Taq* DNA polymerase-generated PCR product. This tag consists of a fluorescent "reporter" and a "silencing" quencher dye at opposite ends of the oligonucleotide. When annealed to the internal sequence of the PCR product, fluorescence from the reporter is quenched because the silencer is in proximity. After completion of each cycle of PCR amplification, the reporter is not incorporated in the product but is cleaved by *Taq* DNA polymerase (because this enzyme also has exonuclease activity). This fluorogenic tag is released, generating a fluorescent signal (Figure 1-3). Real-time PCR detects the number of cycles when amplification of the product is exponential and expresses this as a ratio to standard housekeeping RNA, such as ribosomal RNA or glyceraldehyde-3-phosphate dehydrogenase mRNA. This number can be converted to the number of molecules of mRNA present in the test sample. This technique is used widely to measure minimal residual disease (MRD) or to monitor clearance of *BCR/ABL* transcripts in patients treated with tyrosine kinase inhibitors.

The power of PCR lies in its great sensitivity, but this is also a potential weakness because small amounts of contaminating DNA or RNA from other sources can cause false-positive results. Clinical laboratories that use PCR for critical diagnostic tests require elaborate quality assurance protocols to prevent inappropriate diagnosis. Equally troublesome are false-negative results that result from inappropriate primer design, degraded RNA, or inappropriate temperature parameters for the annealing of primers.

The amplified sequence of interest then can be rapidly evaluated for presence of mutation(s) by direct sequencing, restriction enzyme digestion (if a suitable enzyme that discriminates between mutant and wild-type alleles is available), allele-specific PCR (discussed later in this chapter), or other techniques.

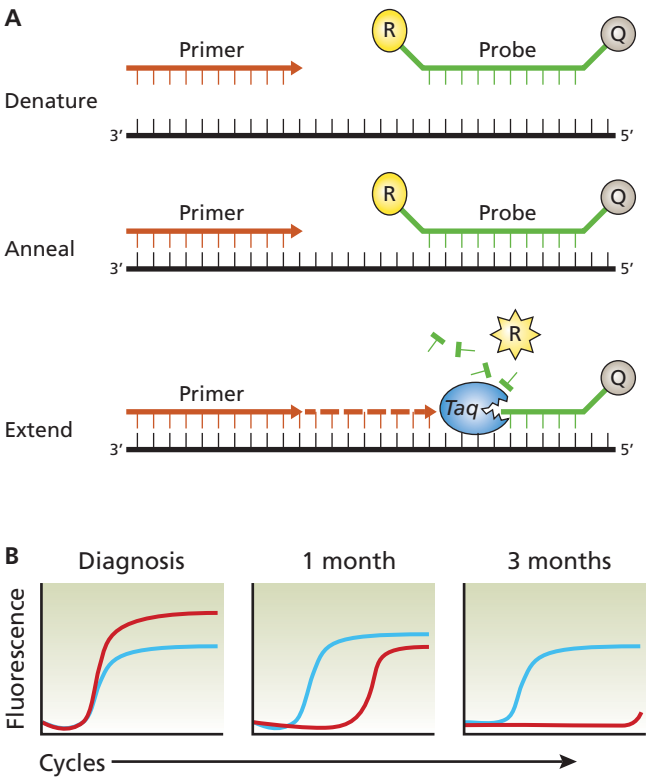


Figure 1-3 Real-time PCR. (A) Sample DNA (or cDNA) is denatured, and target-specific primers are annealed to begin the PCR amplification (shown for one strand). An oligonucleotide probe complementary to a sequence within the PCR product is included in the reaction. The probe contains a fluorophore (R) covalently attached to the 5′ end and a quencher (Q) at the 3′ end. As the *Taq* polymerase extends the nascent strand, its 5′-to-3′ exonuclease activity degrades the probe, releasing the fluorophore from the quencher and allowing the fluorophore to fluoresce. (B) An example of this fluorescent readout is shown, which depicts the relative fluorescence intensity from amplification of the *BCR-ABL* fusion transcript (red line) to an endogenous control transcript (blue line) in a patient with chronic myelogenous leukemia before and after treatment with a tyrosine kinase inhibitor. The cycle number at which fluorescence crosses a threshold is inversely proportional to the amount of template DNA or cDNA. Although the control template is consistently detected throughout therapy, the *BCR-ABL* transcript abundance is lower at 1 month (higher cycle threshold) and undetectable at 3 months into therapy.

Cytogenetic techniques

Uniform, nonrandom chromosomal abnormalities, termed *clonal abnormalities*, can be detected in malignant cellular populations by metaphase **cytogenetics**, or chromosomal analysis. Conventional cytogenetic techniques can detect numeric chromosomal abnormalities (too many or too few chromosomes), as well as deletion or translocation of relatively large chromosomal fragments among chromosomes. Certain chromosomal translocations are considered pathognomonic of specific diseases, such as the t(15;17) in acute promyelocytic leukemia. Normally, chromosomes cannot be seen with a light microscope, but during cell division they become condensed and can be analyzed. To collect cells with their chromosomes in this condensed state, bone marrow or tumor tissue may be briefly maintained in culture and then exposed to a mitotic inhibitor, which blocks formation of the spindle and arrests cell division at the metaphase stage. Thus, cytogenetic studies require dividing cells.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

Conventional cytogenetic studies have several limitations. First, these studies require active cell division, which induces sampling bias (measures dividing cells from a sample) and may not be feasible for some clinical samples. Second, the technique is insensitive to submicroscopic abnormalities. Finally, because only a very small number of cells are analyzed, usually 20 metaphases, the technique cannot be used to determine minimal residual disease.

Fluorescence in situ hybridization (FISH) studies complement conventional cytogenetic analysis by adding convenience, specificity, and sensitivity. This technique applies the principles of complementary DNA hybridization. A specific single-stranded DNA probe corresponding to a gene or chromosomal region of interest is labeled for fluorescent detection. One or more probes are then incubated with the fixed cellular sample and examined by fluorescence microscopy. FISH probes have been developed that can identify specific disease-defining translocations, such as the t(15;17) that characterizes acute promyelocytic leukemia. A probe corresponding to the *PML* gene on chromosome 15 is labeled with a fluorescent marker, such as rhodamine, which is red. Another fluorescent marker, such as fluorescein, which is green, is linked to a probe corresponding to the *RARα* gene on chromosome 17. When the t(15;17) chromosomal translocation is present, the two genes are juxtaposed, the two probes are in proximity, and the fluorescent signals merge to generate a yellow signal. The specificity of FISH is highly dependent on the probes that are used. Numeric abnormalities, such as monosomy and trisomy, may be identified using centromere-specific probes.

The major advantage of FISH is that it can analyze known cytogenetic abnormalities in nondividing cells (interphase nuclei); thus, peripheral blood slides can be directly processed. FISH studies are most useful when assessing for the presence of specific molecular abnormalities associated with a particular clinical syndrome or tumor type and are approximately 1 order of magnitude more sensitive than morphology and conventional cytogenetic studies in detecting residual disease. FISH panels are now available to detect recurrent genetic changes in leukemias, lymphomas, and multiple myeloma. These panels are particularly useful in predicting prognosis when conventional cytogenetic studies are noninformative.

Since their introduction nearly 30 years ago, FISH techniques have evolved rapidly for use in hematologic disorders. For example, double-fusion FISH (D-FISH) uses differentially labeled large probes that each span one of the two **translocation breakpoints**. This allows simultaneous visualization of both fusion products and reduces false-negative results. Another technique known as break-apart FISH uses differentially labeled probes targeting the regions flanking the breakpoint. Thus, in normal cells, the signals appear fused but they split upon translocation. This technique has been used to detect *MYC* translocations in Burkitt lymphoma and *CCND1* translocations in mantle cell lymphoma. Labeling probes with unique combinations of fluorophores in multiplex FISH (M-FISH) not only has permitted simultaneous detection of every chromosome but also now has been used to analyze specific chromosomal regions and can detect subtle rearrangements.

Array-based techniques

DNA **microarrays** are composed of oligonucleotide probes spanning sites of known **single-nucleotide polymorphisms (SNPs)**. Fluorescently labeled single-stranded DNA from a test sample is hybridized on the array to determine, for a specific region in the genome, which DNA sequence undergoes complementary base pairing with the sample. The pattern of hybridization signals is analyzed using computer software, providing a detailed profile of genetic variation specific to an individual's DNA. With current technology, a single microarray has sufficient density to analyze variation at >1 million polymorphic sites. These data can be analyzed in several ways. First, the genotypes at each site can be used in genome-wide association studies in which the allele frequencies at each SNP are compared in disease cases and unaffected controls. Second, the intensity of fluorescent signals from multiple adjacent sites can be used to infer changes in the abundance of DNA across the genome. Changes in DNA content may include inherited **copy number variants** or somatically acquired deletions and amplifications present in tumor samples. Finally, long stretches of homozygosity that reflect acquired **partial uniparental disomy**, a recurrent abnormality present in a variety of myeloid malignancies, can be identified. In a variation of this technique, the relative abundance of methylated versus unmethylated DNA can be detected in samples by pretreating DNA with chemicals (eg, bisulfite) that convert methylated cytosine bases before hybridization on an array.

In **comparative genomic hybridization (CGH)**, DNA extracted from a test sample (eg, tumor) and a matched normal control (eg, buccal wash) is differentially labeled and hybridized to a microarray composed of oligonucleotide probes. The ratio of test to control fluorescence is quantified using digital image analysis. Similar to SNP arrays, amplifications in the test DNA are identified as regions of increased fluorescence ratio, and losses are identified as areas of decreased ratio. In array CGH, resolution of the analysis is restricted by probe size and

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

the density of probes on the array. These and other techniques permit high-resolution, genome-wide detection of genomic copy number changes. Careful analysis of AML genomes using these approaches has revealed few somatic copy number changes that are not detectable by routine cytogenetics. In contrast, ALL genomes are characterized by recurring copy number alterations, frequently involving the loss of the genes required for normal lymphoid development (eg, *PAX5*, *IKZF1*).

RNA expression arrays allow comprehensive characterization of the gene expression patterns within the cells of interest, referred to as a **gene expression profile**. This technique has been used to classify disease, predict response to therapy, and dissect pathways of disease pathogenesis. To perform these assays, mRNA is extracted from samples and double-stranded cDNA is synthesized from the RNA template. Then, biotinylated complementary RNA (cRNA) is generated from the cDNA template by in vitro transcription using biotin-labeled nucleotides. The biotinylated cRNA is fragmented and incubated with probes in a solution or hybridized to a microarray. Hybridization is then detected using a streptavidin-phycoerythrin stain, and the fluorescence intensity of each feature of the array is quantified.

Two main computational approaches have been used to analyze microarray data: unsupervised and supervised learning. Unsupervised learning methods cluster samples based on gene expression similarities without a priori knowledge of class labels. Hierarchical clustering and self-organizing maps are two commonly used algorithms of unsupervised learning. One potential application of unsupervised learning is for discovery of previously unrecognized disease subtypes. The strength of this method is that it provides an unbiased approach to identifying classes within a data set. A weakness is that these data sets are complex, and the structure uncovered by clustering may not reflect the underlying biology of interest. The second computational approach, supervised learning, uses known class labels to create a model for class prediction. For example, a training data set is used to create an expression profile for tumor samples from patients with "cured" versus "relapsed" disease. These profiles then are applied to an independent data set to validate the ability to make the prognostic distinction. In either method, it is important to demonstrate statistical significance and ensure that the tested samples are compared with the appropriate controls.

Once differentially expressed genes are defined, then it is also often helpful to next determine whether the differentially expressed genes that are identified belong to specific pathways of known biological significance. This is commonly done through the use of the GO (Gene Ontology) or KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway analysis database. Also, a statistical methodology known as GSEA (gene set enrichment analysis) is commonly used to determine if the differentially expressed genes are statistically enriched in previously published and defined gene sets publicly deposited in prior gene expression data sets.

The main limitation of microarray expression technology is that it analyzes only mRNA abundance. It does not reveal important translational and posttranslational modifications and protein–protein interactions. Purity of the cell population is also essential for these analyses, and one must ensure that the control and analyzed cells are homogeneous, of the same cell type, and at comparable stages of differentiation.

Sequence-based studies

Analysis of DNA sequence variation by conventional techniques (eg, Sanger sequencing) is being replaced rapidly by a variety of novel high-throughput technologies (collectively termed **next-generation [next-gen] sequencing**). These developments have greatly accelerated the pace and lowered the cost of large-scale sequence production. At the core of each of these technologies is the preparation of DNA fragment libraries, which are then clonally amplified and sequenced by synthesis in multiple parallel reactions (Figure 1-4). Sequencing both ends of the DNA templates ("paired-end reads") improves the efficiency of data production and facilitates the identification of insertions, deletions, and translocations. With these approaches, the search for inherited and somatic mutations associated with hematologic malignancies and congenital blood disorders has evolved from a candidate gene approach to unbiased surveys of all coding and noncoding regions of the genome.

Genome sequencing

The DNA libraries used for these sequencing reactions can be prepared from whole genomes or from selected regions of interest. For example, the regions of the genome that encode proteins (the **exome**) can be enriched by hybridizing DNA to oligonucleotide probes before library construction. Exome sequencing is preferred for many studies (compared with whole genome sequencing) because the cost of sequence production is lower and the interpretation of sequence variants in protein-coding genes is more tractable. A limitation of exome sequencing is that it cannot detect structural variants (such as deletions, amplifications, and rearrangements). In addition, the use of DNA from tissue uninvolved in the disease process is very important for whole exome or genome sequencing

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

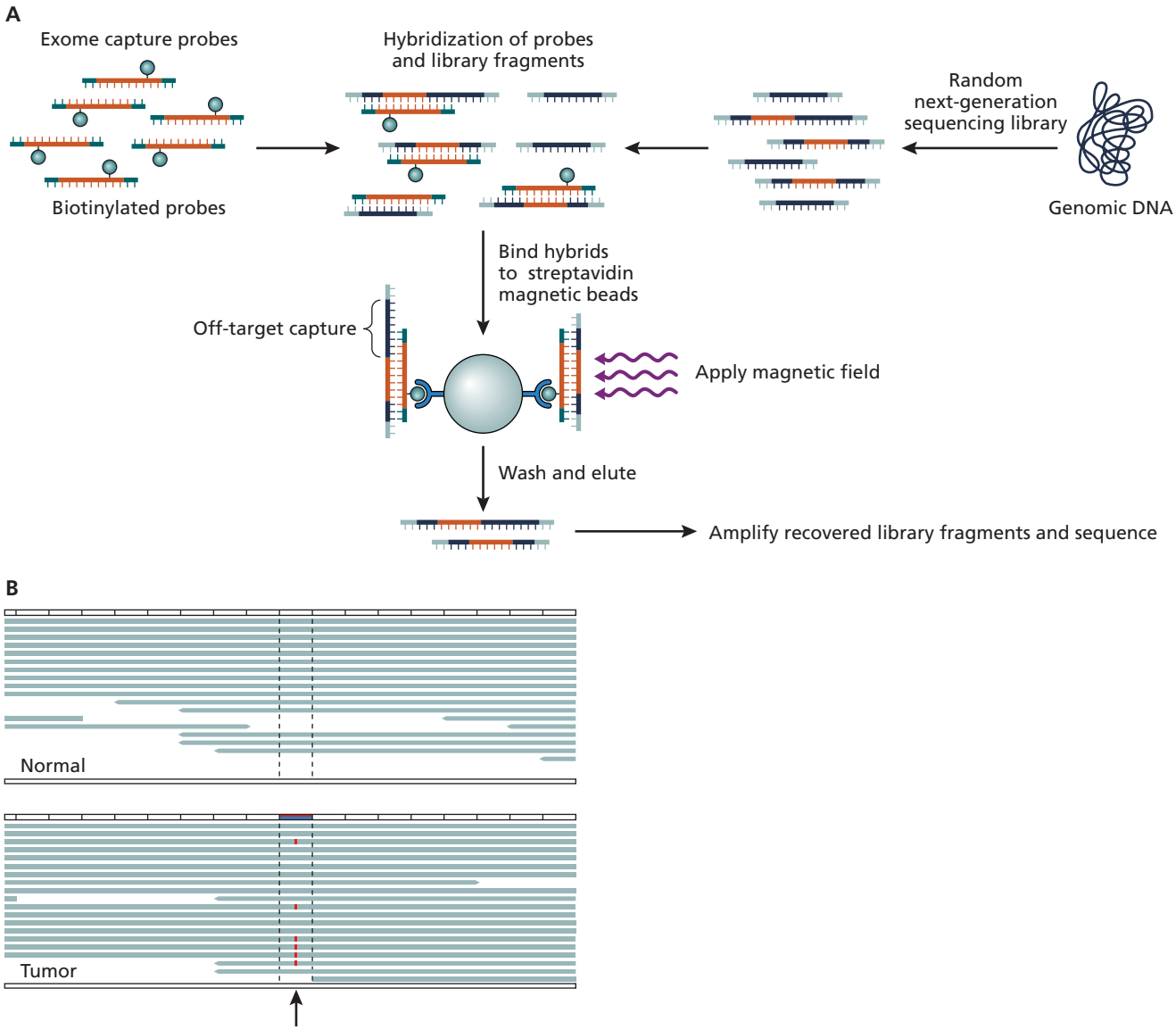


Figure 1-4 Next-gen sequencing. (A) A library of genomic DNA fragments is shown at the upper right. This library can be used directly for sequencing (whole genome) or hybridized with probes (shown in upper left) to enrich for targets in candidate genes or all coding genes (whole exome). Reproduced from Mardis E, *Nat Rev Gastroenterol Hepatol*. 2012;9:477–486, with permission. (B) Sequence reads from a tumor sample (bottom panel) and matched normal tissue (upper panel) visualized in the Integrated Genome Viewer (Robinson JT et al, *Nat Biotechnol*. 2011;29:24–26). Each gray bar represents a single 75- to 100-bp sequence read. A nucleotide substitution shown in red at the position indicated by the arrow is present in ~50% of tumor reads, suggesting that a heterozygous somatic mutation is present in the tumor sample.

to identify potential somatic alterations specific for the disease tissue, given the large number of sequence variants that are often generated by sequencing the entire exome or genome. Also, some protein-coding genes are not yet annotated and therefore are excluded from commercially available reagents used to capture the target DNA. These assays can be restricted further to panels of genes by first amplifying the genes of interest (by PCR) or by hybridizing the DNA to oligonucleotide probes covering the region of interest, followed by next-gen sequencing. This is an efficient and cost-effective approach to interrogate a number of targets in parallel from a single sample (eg, genes that are recurrently mutated in hematologic malignancies).

Transcriptome analysis

RNA templates can be used to generate DNA libraries for sequencing (**RNA-Seq**), allowing for the

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

quantification of RNA abundance and for the detection of **chimeric** RNAs or alternatively spliced products. A specialized type of RNA-Seq, in which individual cells are tagged prior to performing RNA-Seq (single-cell RNA-Seq) offers transcriptome analysis at single cells level. More details follow.

Chromatin and long-range interaction analysis

Further modifications of the workflow allow for detection of chromatin marks across the genome (eg, transcription factor binding sites, histone modifications) by using antibodies to immunoprecipitate the region of interest, followed by next-gen sequencing (**ChIP-Seq**). Next-generation sequencing–based assays can also be used to identify various epigenetic marks. For example, bisulfite sequencing is used to quantify the extent of cytosine methylation across the genome. In this method, sodium bisulfite exposure converts unmethylated cytosines to uracil, leaving methylated cytosines unaffected. Multiple approaches to determine chromatin accessibility and long-range protein–protein interactions have been developed using next-gen sequencing. Assay for transposase-accessible chromatin using sequencing (**ATAC-seq**) utilizes a mutated hyperactive transposase, an enzyme that catalyzes the movement of transposon DNA elements to other parts in the genome. The high activity of the mutant enzyme allows for highly efficient cutting of exposed DNA and simultaneous ligation of adapters. Adapter-ligated DNA fragments are then isolated, amplified by PCR and used for next-gen sequencing. By isolating exposed DNA, the technique identifies areas of the genome where the chromatin is accessible, an indicator of genomic regions free of nucleosomes. Thus, enrichment of sequences indicates absence of DNA-binding proteins or nucleosomes in the region. These regions can be further categorized into regulatory elements—such as promoters, enhancers, and insulators—by integrating further genomic and epigenomic data, including information about histone modifications or evidence for active transcription. Chromosome conformation capture techniques are a set of molecular biology methods used to analyze the spatial organization of chromatin in a cell. These methods quantify the number of interactions between genomic loci that are nearby in three-dimensional space but may be separated by long distances in the linear genome. Such interactions may result from biological functions, such as promoter–enhancer interactions. **Hi-C** is one of these techniques that quantifies interactions between all possible pairs of fragments simultaneously. Briefly, cell genomes are cross-linked with a fixative to "freeze" interactions between genomic loci. The genome is then cut into fragments using restriction enzymes and a ligase is added to cross-link interacting fragments. Interacting loci fragments then undergo library preparation followed by high-throughput sequencing.

Although the cost of sequence production has fallen dramatically in recent years, the storage, analysis, and interpretation of these large data sets still pose significant challenges.

Methods to study protein abundance

Proteins are the effectors of most cellular functions. Genetic defects perturb normal cellular functions because they result in changes in the level or function of the proteins they encode. Characterizing proteins expressed on cell surfaces and within cells is critical for identifying hematopoietic and immune cell subsets and is a cornerstone of diagnosing a wide variety of hematologic malignancies. Evaluation of protein expression abundance for diagnostic and therapeutic purposes is routinely performed using either single-cell analysis such as **flow cytometry** of suspension cells as well as **immunohistochemistry (IHC)** analysis of tissue sections or bulk analysis such as enzyme-linked immunosorbent assay (ELISA), sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), Western blotting, or protein microarrays to detect peptides, proteins, antibodies, and hormones in liquid or tissue material. Many proteins undergo extensive posttranslational modifications that influence their activity and function, including cleavage, chemical modification such as phosphorylation and glycosylation, and interaction with other proteins. These posttranslational events are not revealed by genomic analysis or gene expression profiling. **Proteomics** is the systematic study of the entire complement of proteins derived from a cell population.

IHC, flow cytometry, and ELISA analyses are heavily utilized in clinical diagnostic hematopathology. IHC analysis is the process of visualizing the expression of a protein within a section of tissue through the use of antibodies selective to a specific antigen. Most commonly, the antibody utilized in IHC is conjugated to an enzyme, such as peroxidase, which catalyzes a color-producing reaction allowing visualization of the antibody-antigen interaction. Because IHC is performed on a tissue section, the architecture of the tissue and cellular relationships in tissue is well preserved. In addition, IHC is performed on tissue that has been preserved through the process of fixation, most commonly using paraformaldehyde, allowing IHC analysis on archival tissue. IHC immunophenotyping is routinely used to differentiate subtypes of acute leukemias and lymphomas using panels of defined antibodies.

Similar to IHC, in ELISA, proteins are incubated with an antibody linked to an enzyme where the abundance

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

of the protein is indicated by the extent of the enzymatic activity. Unlike IHC, however, ELISA is performed on any source where protein can be extracted and is routinely used to detect peptides, proteins, antibodies, and hormones in clinical materials.

In contrast to IHC and ELISA, flow cytometry is utilized to characterize protein expression on cells obtained viably in single-cell suspension. In flow cytometry, antibodies conjugated to a fluorescent protein bind to a cell surface (and/or in a cell that has been permeabilized for detection of proteins within a cell) and are passed through an electronic detection apparatus to enumerate the number of cells expressing that protein. In this manner, panels of antibodies with different fluorescent proteins are utilized to characterize numbers of proteins simultaneously on peripheral blood, bone marrow aspirate, and tissue fluid samples. In addition, flow cytometry provides quantitative information and has a level of sensitivity that allows it to be used for testing of minimal residue disease following treatment (described later). For both flow cytometry and IHC, variations of these techniques have been developed recently using antibodies conjugated to metal ions, which allows for the use of much larger number of antibodies simultaneously.

Proteomic analysis relies on complex bioinformatic tools applied to mass spectroscopy data. In general, these techniques require some sort of separation of peptides, usually by liquid chromatography, followed by ionization of the sample and mass spectrometry. In matrix-associated laser desorption/ionization–time of flight mass spectrometry, the time of flight of the ions is detected and used to calculate a mass-to-charge ratio. The spectrum of mass-to-charge ratios present within a sample reflects the protein constituents within the sample. Supervised or unsupervised learning approaches, as described previously, are then used to identify patterns within the data. More recently, protein microarrays have been developed. Analytical protein microarrays are composed of a high density of affinity reagents (eg, antigens, antibodies) that can be used to detect the presence of specific proteins in a mixture. Functional protein microarrays contain a large number of immobilized proteins; these arrays can be used to examine protein-protein, protein-lipid, protein–nucleic acid, and enzyme-substrate interactions. Although all of these technologies hold enormous potential, clinical applications have yet to be realized.

Single-cell molecular profiling

The last decade has seen a number of technological innovations that now facilitate the profiling of molecular entities (DNA, RNA, protein) at the level of individual single cells. A major driving force behind this is the recognition that cellular heterogeneity may well be a ubiquitous feature not just of hematology research, but biomedical research more broadly. Consequently, conventional "bulk" approaches provide average measurements that obscure such cellular heterogeneity, may miss subpopulations altogether, and may provide average values that are not representative of any single cell present in the sample. To make single-cell molecular profiling a reality, technical innovation in two areas was critical, namely the miniaturization and parallelization of assays. This in turn enables the profiling of the small amounts of material present in a single cell, and at the same time provides this information for large numbers of single cells.

The most advanced modality for single-cell molecular profiling is RNA, where single-cell transcriptomics by single-cell RNA-Seq (scRNA-Seq) is now commonly performed in hematology research settings. Early protocols were plate-based, where individual cells were deposited in a 96- or 384-well plate and processed with reverse transcriptase followed by PCR-based cDNA amplification. Incorporation of barcode sequences permits subsequent pooling of samples, to permit highly efficient multiplexed sequencing of thousands or tens of thousands of cells at once. The more widely used protocols now take advantage of droplet technology, whereby individual cells are captured within droplets that serve as the reaction chambers for the initial steps of reverse transcription and incorporation of sequencing barcodes. Commercial platforms exist for single-cell profiling of RNA, DNA, methylated DNA, open chromatin, cell surface protein markers, as well as an ever-expanding array of commercial and non-commercial platforms for so-called single-cell multiomics, whereby measurements for more than one modality are obtained for each single cell. One of the single-cell multiome methods most relevant for hematology is CITE-seq, which provides full transcriptomic information coupled with quantification of dozens to hundreds of proteins, from thousands of individual single cells. Some of the biggest current challenges in single-cell technology are concerned with computational analysis of the complex data sets that contain detailed molecular information for thousands of individual single cells. It is nevertheless anticipated that the use of single-cell profiling will become ever more widespread and, for a number of both research and clinical/diagnostic applications, is likely to replace the more traditional "bulk" analysis methods eventually.

Animal models

Analysis of both inherited and acquired diseases by reverse genetics has resulted in the identification of many

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

disease-related genes for which the function is unknown. Once a disease-related gene has been identified, either by **linkage mapping** (eg, the gene for cystic fibrosis) or by identifying rearranged genes (eg, the *BCR* gene at the breakpoint of the Philadelphia chromosome), the challenge lies in identifying the function of the protein encoded by that gene and characterizing how changes in the gene can contribute to the disease phenotype. Understanding the role of these genes and their encoded proteins has been aided greatly by the development of techniques to alter or introduce these genes in mice using recombinant DNA technology.

Mice generated via random genomic insertion of a transgene

Mice can be produced that express an exogenous gene and thereby provide an in vivo model of the gene's function. Linearized DNA is injected into a fertilized mouse oocyte pronucleus, and the oocyte is then reimplanted into a pseudopregnant mouse. The resultant **transgenic mice** then can be analyzed for phenotypes induced by the exogenous gene. Placing the gene under the control of a strong **constitutive promoter** creates a "constitutive transgenic animal" in which the transgene is always expressed. A number of inducible promoter-enhancer elements can be used to control transgene expression. Such "inducible transgenic animals" allow the use of tetracycline or tamoxifen for instance to turn on or off genes of interest during embryonic life or adulthood. In addition to controlling the timing of transgene expression, the tissue distribution of this expression can also be controlled via the use of tissue-specific promoter-enhancer elements, creating a "tissue-specific transgenic animal." A unique type of "tissue-specific transgenic animals" are generated when the transgene used can be detected by chemical, immunologic, or functional means creating thus a "reporter animal." For example, the promoter region of a gene can be joined to the green fluorescent protein cDNA, and expression of this reporter can be assessed in various tissues in the resultant transgenic mouse. Use of such a reporter gene will show the normal distribution and timing of the expression of the gene from which the promoter elements are derived.

Mice generated via homologous recombination

The transgenic mice mentioned previously ordinarily contain multiple copies of exogenous genes that have inserted randomly into the genome of the recipient and thus may not mimic physiologic levels or spatiotemporal expression of the gene. In contrast, the endogenous genetic locus of a gene can be manipulated in totipotent embryonic stem (ES) cells by targeted recombination between the locus and a plasmid carrying an altered version of that gene that changes or disrupts its function. If a plasmid contains that altered gene with enough flanking DNA identical to that of the normal gene locus, **homologous recombination** will occur at a low rate; however, cells undergoing the desired recombination can be enriched by including a selection marker in the plasmid, such as the neomycin resistance gene. The correctly targeted ES cell is then introduced into the blastocyst of a developing embryo. The resultant animals will be chimeric, in that only some of the cells in the animal will contain the targeted gene. If the new gene becomes part of the germline, offspring can be bred to yield mice carrying the mutation in all cells. **Knockout mice** (homozygous for a null allele) can illuminate the function of the targeted gene by analyzing the phenotype of mice that lack the gene product. Similar approaches can be used to replace a normal mouse gene in ES cells with a version containing a point mutation, deletion, or other genetic variant to model abnormalities detected in patients with hematologic disorders.

Many genes of interest participate in pathways that are vital for viability or fertility; thus, constitutive knockout mice cannot be generated. Conditional gene modification using Cre-*loxP* technology allows the gene of interest to be altered in specific tissues or at specific times during development or postnatal life. This is accomplished by inserting the altered gene with flanking DNA containing *loxP* sites. If mice with paired *loxP* sites integrated into their genome are bred with a second strain of mice that express an enzyme called Cre recombinase, recombination will take place between the *loxP* sites, removing or rearranging the desired portion of the gene. Furthermore, expression of Cre recombinase can be regulated in a tissue-specific manner by using an appropriate promoter or in a temporally restricted manner by using a promoter that is induced by treatment of the mice with a drug (such as tetracycline). The use of transgenic, knockout, and conditional knockout mice has been invaluable in elucidating the function of large numbers of genes implicated in the pathogenesis of both inherited and acquired diseases.

Using similar homologous recombination technology, transgenes can be inserted into the locus of an endogenous gene generating **knockin mice**. These transgenes, which can be reporter genes or sometimes point mutations of the endogenous genes will intimately reproduce the timing and spatial expression of the endogenous gene.

Disease-in-a-dish models

Transgenic technology, however, is laborious, time consuming, and expensive. Some of these disadvantages can

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

be circumvented by using rapidly reproducing and inexpensive organisms, such as zebrafish or yeast. Like transgenic mice, however, these models may not recapitulate human-specific pathophysiology. Newer technology using dedifferentiated somatic cells reprogrammed to become totipotent cells may overcome some of these obstacles. These cells, called **induced pluripotent stem cells (IPSCs)**, are produced by reprogramming adult somatic cells to become embryonic-like cells, which, in turn, can be further differentiated along specific lineages. The concrete demonstration that iPSCs may be used to treat disease was replacement of the sickle globin gene with a normal β-globin gene in mice. Corrected iPSCs from sickle mice were differentiated into hematopoietic progenitors in vitro, and these cells were transplanted into irradiated sickle mice recipients. Erythroid cells derived from these progenitors synthesized high levels of human hemoglobin A and corrected the sickle cell disease phenotype. Human iPSCs have been produced and hold great promise as research tools and possibly as a source of tissue replacement.

Given the time required for conventional gene targeting using homologous recombination, there has been great interest in the recent development of genome editing using **zinc finger nucleases, transcription activator–like effector nucleases (TALENs)**, and **CRISPR/Cas (clustered regularly interspaced short palindromic repeat/Cas-based RNA-guided DNA endonucleases)**. Each of these techniques makes use of a nuclease that induces DNA breaks and then stimulates DNA repair in a way that allows for creation of specific mutations and/or inclusions of novel sequences of DNA. These techniques have allowed for rapid generation of knockout and knockin mice in embryonic or somatic stem cells to rapidly create genetically engineered animal models.

Clinical applications of DNA technology in hematology

Molecular biology has revolutionized the understanding of molecular pathogenesis of disease in ways that have profoundly affected the diagnostic armamentarium of the hematologist. Several examples of how molecular studies are used for diagnosis and clinical decision-making in hematology are described in this section.

Applications to germline (inherited) mutations

Hemoglobinopathies and thalassemias

One of the best examples of the use of molecular techniques in benign hematology is in the diagnosis of hemoglobinopathies and thalassemia. Although the most common hemoglobin variants (ie, Hb S, Hb C, Hb D) are typically diagnosed using nonmolecular methods, such as high-performance liquid chromatography or protein electrophoresis, molecular testing can be useful in several settings, including the characterization of uncommon variants, family screening studies, and prenatal diagnosis. Hemoglobin variants may be detected by a variety of techniques, including PCR, using allele-specific primers designed to detect specific mutations or sequencing studies of the *HBA1/A2* and *HBB* loci. Molecular techniques are particularly valuable in the diagnosis of α-thalassemia, which usually is caused by one of several variably sized deletions that result in the loss of one or both *HBA* genes in the α-globin locus. In the neonatal period, α-thalassemia may be recognized by the presence of Hb Barts (4 tetramers) on electrophoresis or high-performance liquid chromatography, but laboratory diagnosis after the neonatal period requires molecular techniques. Deletions of the α-globin locus can be detected by gap-PCR, which uses PCR primers that bind to either side of a deletion breakpoint. In the absence of the corresponding deletion, the primers are too far apart to yield an amplifiable product. When a deletion is present, however, an abnormal amplicon is detected.

Pharmacogenomics

Pharmacogenomics is the study of how inherited genetic variation affects the body's response to drugs. The term comes from the words *pharmacology* and *genomics* and is thus the intersection of both disciplines. For instance, homozygous germline **polymorphisms** in the thiopurine methyltransferase (TPMT) gene result in loss of functional protein and predispose ALL patients to severe hematologic toxicity unless the dose of mercaptopurine is reduced by 90% to 95% of normal. Heterozygote individuals also require dose reductions to a lesser extent than homozygotes. PCR-based studies may be performed to identify the presence of alleles associated with decreased TPMT function.

In addition to hemoglobinopathies (see Chapter 7) and pharmacogenomics, germline mutations are recognized to play an increasing role in the pathophysiology of various hematological conditions, including hemostasis (Chapter 10) and thrombosis (Chapter 9), bone marrow failure syndromes (Chapter 15), and transfusion medicine (Chapter 13), including red cell and platelet genomics.

Applications to somatic (acquired) molecular abnormalities

The power of molecular biology to provide important insights into the basic biology of disease is perhaps most

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

dramatically shown by the evolving concepts of malignancy. Several examples of how molecular techniques have enhanced our understanding of the pathogenesis of hematologic malignancies, as well as their diagnosis and treatment, are provided in the following sections.

Gene rearrangement studies in lymphoproliferative disease: T-cell and B-cell rearrangements

During the development of a mature lymphoid cell from an undifferentiated stem cell, somatic rearrangements of the immunoglobulin and T-cell receptor loci take place, resulting in an extensive repertoire of composite genes that creates immense immunoglobulin and T-cell diversity. These somatic rearrangements result in deletion of intervening DNA sequences between gene segments in the immunoglobulin and T-cell receptor loci. The details of this process in lymphocyte ontogeny are further outlined in Chapter 20.

Rearrangements in immunoglobulin and T-cell receptor genes can be detected by either **Southern blotting** or PCR; however, PCR-based approaches using standardized, comprehensive primer sets such as those developed by the EuroClonality consortium (so called BIOMED-2 primers) are now preferentially used in the clinical setting because they are more rapid, require less DNA, and can be performed on archived formalin-fixed, paraffin-embedded (FFPE) tissue. PCR-based techniques targeting *IGH* and *IGK* loci for B-cell rearrangements and *TCRG* and *TCRB* for T-cell rearrangements are used to confirm the presence of clonal lymphocyte populations in the peripheral blood, such as in T-cell large granular lymphocyte disorders, and also are powerful ancillary techniques for hematopathologists in the diagnosis of lymphoproliferative disorders from FFPE tissue. Despite their power, molecular clonality studies should be carefully interpreted in the context of the clinical, morphologic, and immunophenotypic diagnosis. Clonal proliferations may occur in some reactive conditions as well as in malignant neoplasms. For example, clonal T-cell populations may be detected in the setting of viral infections, such as with Epstein-Barr virus or cytomegalovirus, and clonal B-cell populations may be detected in some benign lymphoid proliferations, such as marked follicular hyperplasia. Furthermore, false-positive PCR results may occur in several circumstances; for example, when very small tissue samples are used, a few reactive T cells in the sample might result in the appearance of oligoclonal bands. False-negative results may occur as a result of tissue sampling, poor PCR amplification, or lack of detection of specific rearrangements using standardized primer sets.

Identification of cryptic translocations in pediatric leukemia: prognostic significance

Several fusion events in pediatric acute leukemia that carry prognostic significance are not detected by standard cytogenetic techniques and require molecular testing for identification. Examples include the *ETV6-RUNX1* fusion that is present in ~20% of children with pre–B-cell ALL and confers a favorable prognosis, as well as high-risk fusion events found in AML, such as *MLL-AF10*, *CBFA2T3-GLIS2*, *NUP98-KDM5A*, and *NUP98-NSD1*, among others. These can be detected by FISH, RT-PCR, or next-gen sequencing approaches such as RNA-Seq or whole genome sequencing. The identification of these lesions is important because they impact the intensity of chemotherapy treatment given to the patient up front, as well as the recommendation for stem cell transplant in first remission.

Prognostically significant mutations in normal karyotype acute myeloid leukemia

Up to 40% of AML cases have no chromosomal abnormalities visible by conventional karyotyping. The prognosis in these cases can be further refined by molecular testing for mutations in various recurrently mutated genes, including *NPM1*, *FLT3*, *CEBPA*, *DNMT3A*, and others. *NPM1* mutations, usually a 4-bp insertion in exon 12, are found in approximately 35% of cases of AML. *FLT3* mutations include variably sized duplications (internal tandem duplications) or point mutations in the kinase domain and are found in approximately one-third of AML cases. Mutations in *CEBPA* are diverse and can be found in approximately 10% of AML cases. Cases of AML with a normal karyotype and a mutant *NPM1*/wild-type *FLT3* genotype or harboring biallelic *CEBPA* mutations are associated with a favorable prognosis. Furthermore, studies have suggested that AML with mutated *NPM1* or mutated *CEBPA* each represent distinct clinicopathologic entities. Using multivariate analysis, mutations in *DNMT3A* have emerged as powerful predictors of poor prognosis in AML. Currently, a limited number of genes are routinely tested in AML patients by conventional (Sanger) sequencing or PCR assays. The use of next-gen sequencing panels has allowed for improved prognostic assessment and treatment selection based on testing a larger number of recurring mutations in myeloid neoplasms, including AML, MDSs, and myeloproliferative neoplasms.

Minimal residual disease monitoring

The development of PCR has markedly increased the sensitivity of tests available for the monitoring of MRD in myeloid and lymphoid neoplasms. With the availability

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavriilaki on 10 December 2022

of real-time PCR, the relative abundance of specific transcripts can now be monitored to assess trends of increase or decrease over time. For example, real-time quantitative RT-PCR is used routinely in CML to risk-stratify patients based on transcript quantity rather than simply the presence or absence of a transcript (as discussed previously in this chapter). The accuracy and reliability of real-time quantitative PCR as a measure of *BCR-ABL1* transcript level depends on the quality control procedures carried out by the laboratory. Normalization of the results to an appropriate control gene is required to compensate for variations in RNA quality and the efficiency of the **reverse transcriptase** reaction. *BCR* and *ABL1* have been used as control genes, and both seem to be suitable because they are expressed at low levels and have similar stability to *BCR-ABL1*. The introduction of internationally recognized reference standards now has allowed for reporting of results on the International Scale, which allows for direct comparisons of results among laboratories, even those using different control genes. A major molecular response to imatinib has been defined as a 3-log reduction in *BCR-ABL1* transcripts (*BCR-ABL1*/reference gene) compared with a standardized baseline obtained from patients with untreated newly diagnosed CML, corresponding to 0.1% on the International Scale.

In similar fashion, PCR analysis of immunoglobulin or T-cell receptor gene rearrangements allow the detection of residual disease in the blood or bone marrow of patients who have undergone treatment of a lymphoid malignancy. Because each gene rearrangement is unique, however, the PCR detection of gene rearrangements at this level of sensitivity is labor intensive. PCR of tumor tissue is performed using primers based on consensus sequences shared by the variable and joining regions of the appropriate locus (immunoglobulin or T-cell receptor genes). The specific rearrangement must then be sequenced so that an oligonucleotide specific to the unique rearrangement in that patient's tumor can be synthesized. PCR can then be performed using this **allele-specific oligonucleotide**, with adequate sensitivity to detect 1 in 10^6 cells. As these assays become increasingly available, they will play an important role in estimating prognosis and determining eligibility for autologous transplantation and other therapeutic modalities.

In addition to MRD monitoring by molecular monitoring of DNA/cDNA retrieved from hematopoietic cells, flow cytometric analysis to detect residual leukemic cells in peripheral blood or bone marrow has also been heavily used for MRD monitoring in ALL and more recently in pediatric AML (amongst other hematological malignancies). Studies in pediatric T-lineage ALL demonstrate that while molecular techniques are more sensitive than flow cytometry in detecting residual disease, this fails to predict relapse more accurately. Similarly, in pediatric AML, MRD levels of 0.01% by flow cytometry do not identify patients at lower risk for relapse compared with those that achieve less than 0.1% following induction chemotherapy, suggesting that a threshold that predicts relapse exists beyond which higher levels of sensitivity do not provide prognostic relevance. Further efforts to understand the sensitivity of flow cytometric MRD detection relative to molecular techniques are underway. As the two techniques are complementary, and there exist cases that are not suitable for one or the other, both approaches can be used in tandem to provide optimal prognostic information.

Expression profiling: applications to diagnosis and treatment

Gene expression microarray studies have facilitated the classification of lymphomas and outcome prediction for specific patient populations with this disease. For example, expression profiling was used to create a prediction model that identified two categories of patients with diffuse large B-cell lymphoma: germinal center and activated B-cell types, which carry favorable and unfavorable prognoses, respectively. In addition, expression profiling studies have identified genes overexpressed in patients with poor prognosis, some of which may represent potential therapeutic targets. Of note, microarray expression profiling is increasingly superseded by next-generation sequencing–based methods such as RNA-Seq.

Applications to stem cell transplantation

Human leukocyte antigen typing for stem cell transplantation

Molecular techniques have been important to the further understanding of the diversity of human leukocyte antigen (HLA) genotypes. Serologic testing for HLA antigens often identifies broad groups of cross-reactive antigens. Because there is an increased incidence of severe graft-versus-host disease in patients who receive transplantations from serologically compatible but genotypically incompatible unrelated donors, it is important to identify the individual antigens within these cross-reactive groups. Genotypic HLA typing can be achieved by PCR amplification of the HLA locus, followed by hybridization to specific oligonucleotides corresponding to the different alleles within a given cross-reactive group. Such genotyping is much more predictive of successful transplantation and the risk of graft-versus-host disease than serologic study or the mixed lymphocyte assay, and it has supplanted these assays for the identification of optimal donors,

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

especially unrelated donors. Comprehensive genotyping using SNP arrays may improve HLA matching. This is discussed in detail in Chapter 14.

Analysis of bone marrow engraftment

When donor and recipient are of opposite sexes, the assessment of donor engraftment is based on conventional cytogenetics and is relatively straightforward. When donor and recipient are of the same sex, RFLP analysis of donor and recipient bone marrow allows the detection of polymorphic markers to distinguish DNA from the donor and recipient. After transplantation, RFLP analysis of recipient peripheral blood cells then can be used to document engraftment, chimerism, graft failure, and disease relapse. In most centers, PCR amplification and genotyping of short tandem repeat or variable number tandem repeat sequences that are polymorphic between donor and recipient pairs are now used to assess chimerism.

Applications to novel therapies

Functional genomics for discovery and diagnosis

The recognition that abnormal expression of oncogene products plays a role in malignancy has led to the proposal that suppression of that expression might reverse the neoplastic phenotype. One way of blocking mRNA expression is through the use of **antisense oligonucleotides**. These are short pieces of single-stranded DNA or RNA, 17 to 20 bases long, which are synthesized with a sequence complementary to the transcription or the translation initiation site in the mRNA. These short single-stranded species enter the cell freely, where they complex to the mRNA through the complementary sequence. Investigation of the mechanism of action of antisense oligonucleotides led to the discovery that naturally occurring double-stranded RNA molecules suppress gene expression better than antisense sequences and helped to unravel the mechanism of RNA interference. RNA interference has significant advantages over antisense therapy in that much lower concentrations are required. Numerous studies are under way in hematologic diseases; however, methods for delivery of siRNAs are still far from perfect. In one study, adult stem cells from sickle cell patients were infected with a viral vector carrying a therapeutic γ-globin gene harboring an embedded siRNA precursor specific for sickle β-globin. The newly formed red blood cells (RBCs) made normal hemoglobin and suppressed production of sickle β-globin. In another study, a retroviral system for stable expression of siRNA directed to the unique fusion junction sequence of *ETV6-PDGFRB* resulted in profound inhibition of *ETV6-PDGFRB* expression and inhibited proliferation of *ETV6-PDGFRB*–transformed cells. When applied to mice, this strategy slowed tumor development and death in mice injected with these cells compared with cells not containing the siRNA. Stable siRNA expression sensitized transformed cells to the *PDGFRB* inhibitor imatinib, suggesting that stable expression of siRNAs, which target oncogenic fusion genes, may potentiate the effects of conventional therapy for hematologic malignancies.

The COVID-19 pandemic resulted in a rapid acceleration of the clinical application of RNA therapeutics, initially in the form of novel RNA-based vaccines. It is anticipated that the accompanying technology development will also accelerate the use of RNA-based therapies in other settings, including novel RNA vaccines to target hematological malignancies.

Gene therapy

The application of gene therapy to genetic hematologic disorders has long been an attractive concept. In most cases, this involves insertion of normal genes into autologous hematopoietic stem cells with subsequent transplantation back into the patient. Candidate hematologic diseases for such therapy include hemophilia, sickle cell disease, thalassemia, and severe combined immune deficiency syndrome. Rapid advances in technology for the separation of hematopoietic stem cells and techniques of gene transfer into those cells have advanced efforts toward this goal, and many clinical trials have been completed. Although significant methodologic hurdles remain, research in this field continues to move forward. It should be recognized, however, that correction of such diseases as hemophilia, sickle cell disease, and thalassemia requires efficient gene transfer to a large number of hematopoietic stem cells with high levels of expression of the β-globin gene in erythroid precursors. Long-term repopulating stem cells have been relatively resistant to genetic modification; thus, many investigators have focused on gene therapy applications in which low levels of expression could restore patients to health. A major impediment to successful gene therapy has been the lack of gene delivery systems that provide safe, efficient, and durable gene insertion and that can specifically target the cells of interest. An important safety concern with viral vectors that integrate into the host genome is the potential to activate oncogenes or inactivate tumor suppressor genes by insertional mutagenesis. Currently used approaches include retroviral vectors, adenoviral vectors, other viral vectors, and nonviral vectors. One of the more recent successes in the field that has overcome these challenges has been recently reported for hemophilia B. High expression levels of a functional factor IX was found in patients treated with a single injection of adeno-associated

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

viral vector containing a hyperfunctional factor IX variant gene. All participants in the study had sustained factor IX levels one-third of the normal value, with dramatically reduced annual bleeding rates.

CRISPR/Cas9 and genome editing

A revolution in genome editing is centered on CRISPR/Cas9 technology because of its ready applicability coupled with an improved safety profile over many previous methods of genome editing. The first report of CRISPR/Cas9–mediated genome editing was less than a decade ago. Since then, this system has been successfully used not only to generate preclinical models of disease and to perform functional genome screen but most recently for human gene therapy and to manufacture genetically engineered cellular therapy products. The CRISPR/Cas9 system is highly efficient in correcting hematopoietic stem cells from patients with sickle cell disease and β-thalassemia. The corrected cells can then be reinfused into the patients providing lifelong normal erythropoiesis. An alternative approach is to deliver CRISPR/Cas9 vectors to correct germline mutations in vivo. This strategy is currently being tested for patients with hemophilia, transthyretin amyloidosis, and cystic fibrosis. In addition to the gene therapy applications, CRISPR/Cas9 found its use in the manufacturing of cellular therapy products such as chimeric antigen receptor T cells and genetically modified natural killer (NK) cells.

Cellular basis of hematopoiesis

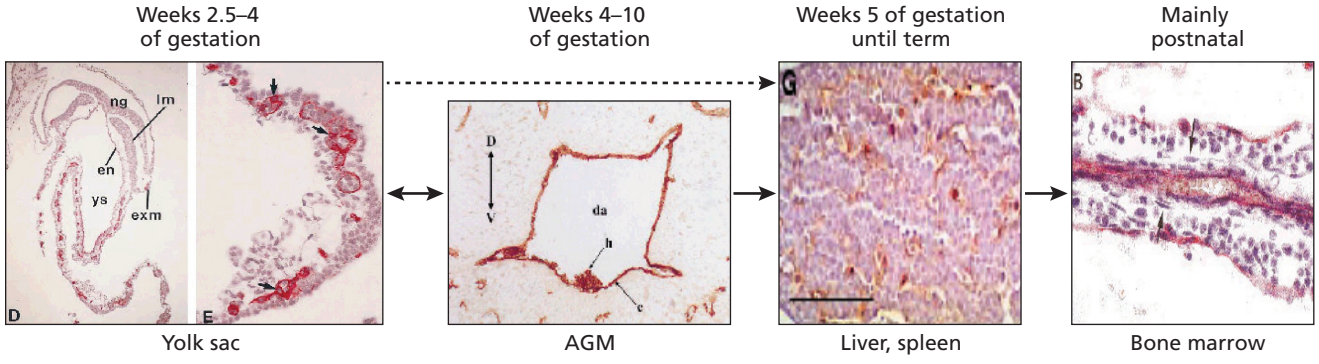
Introduction and historical perspective

Humans produce approximately 300 billion blood cells per day. Hematopoiesis is the process by which blood cells are made by transiting through a hierarchy of hematopoietic stem and progenitor cells (HSPCs). Hematopoietic stem cells (HSCs) are defined by their ability to self-renew as well as differentiate into the progenitors that replenish the entire blood system. Residing at the top of this hierarchy, HSCs are located in a number of embryonic niches, settling in the bone marrow (BM) in adult life. Contrary to previous assumptions, HSCs are a heterogeneous cell population and likely number tens of thousands of cells in adult life, giving rise to hundreds of millions of hierarchically organized, highly heterogeneous progenitor cells, which in turn differentiate into precursor cells and eventually mature effector cells. Although the field of stem/progenitor cell biology has grown dramatically over the past decades, bone marrow transplantation (BMT) has been in routine clinical practice for more than 50 years and is the only routine, widely used example of stem/progenitor cell therapy. Identification of BM HSPCs emerged after it was recognized that survivors from atomic bomb explosions in 1945 died of hematopoietic failure from radiation damage. In the early 1960s, a series of seminal experiments by Till and McCulloch showed that the transfer of BM cells from donor mice into lethally irradiated host mice resulted in the formation of macroscopic colonies (called spleen colony-forming unit [CFU-S] cells) of myeloid, erythroid, and megakaryocytic cells in the spleens of the recipients. These colonies arise from a single implanted cell and were the first demonstration of the existence of a repopulating hematopoietic cell that could also differentiate. This Chapter summarizes the development, differentiation, and localization of HSCs and the hematopoietic niche, as well as aspects of HSC purification and expansion for transplantation.

Hematopoiesis through development and into adult life

Hematopoiesis occurs in waves and at multiple discrete anatomical sites that change through development (Figure 1-5). In humans, like other vertebrates, the initial

Figure 1-5 Changing anatomical locations of hematopoiesis through development. Hematopoiesis is initially detected in the extra embryonic yolk sac, then in the embryo in a region known as the aorta-gonad-mesonephros (AGM), the placenta, the umbilical arteries, and vitelline vessels. It then shifts to the fetal liver and finally to the bone marrow. en, endoderm; exm extraembryonic mesoderm.



Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

wave of hematopoiesis occurs in the extraembryonic yolk sac (YS) blood islands from weeks 3 to 6 of gestation. The YS primarily produces primitive erythroid cells (termed *primitive erythropoiesis*), expressing embryonic globins that deliver oxygen to tissues in the rapidly growing embryo. Additional studies suggest that primitive hematopoiesis produces myeloid and lymphoid cells (eg, macrophages and NK cells). Interestingly, the developmental potential of embryonic hematopoiesis closely resembles hematopoietic cells derived from human embryonic stem cells (hESCs) and iPSCs.

Primitive hematopoiesis is transient and replaced by adult or definitive hematopoiesis that sustains blood production throughout development and postnatal life. Embryonic hematopoietic activity is detected at 4 to 5 weeks of gestation in a region around the ventral wall of the dorsal aorta, called the aorta-gonad mesonephros (AGM). In early development, blood cells arise in close connection with the vascular structures (both in the YS and the dorsal aorta), giving rise to the notion that there may be a common precursor cell population that produces both blood and blood vessel cells called a hemangioblast, or that hematopoiesis arises directly from specialized "hemogenic" endothelial cells, such as those lining the ventral aspect of the dorsal aorta. In mice and other animals, studies have shown that definitive hematopoietic stem cells with serially transplantable activity, together with long-term engraftment capacity, are found in the AGM. It is still a matter of much debate whether true HSCs arise only from the AGM, or also in other intra embryonic sites and/or by colonization from the YS.

HSCs are then detected in the developing fetal liver, spleen, and thymus, from 6 to 22 weeks in humans, where they expand and differentiate into committed progenitor cells. Expansion and differentiation of HSCs allows for development of definitive red cells, myeloid cells, and lymphoid cells (T cells that develop in the thymus and B cells in the marrow). During the later part of this time window, HSCs are detected in the BM. Elucidating the precise origins and route of migration of HSCs in the developing fetus is likely to continue being informed by studies in model organisms that are more experimentally tractable.

Murine HSCs are heterogeneous with respect to lineage output–generating megakaryocyte, myeloid or lymphoid progenitors, or balanced progenitor output. Quantitation of HSC output in mice has been more technically feasible than studies in humans. Using limiting dilution analysis of various tissues, a quantitative and temporal relationship has been modeled in developing mice. AGM and YS HSCs enter the circulation and provide a pool of HSCs recruited to the liver. In whole BM cells, the estimated frequency of HSCs, defined by functional ability for long-term engraftment and multilineage reconstitution, is 3 per 100,000 cells. In the mouse, fetal HSCs show extensive proliferation and tend toward greater lymphoid output. During adult life, HSCs are more quiescent. Functional HSCs are reduced with age, and lymphocyte-producing HSCs diminish relative to myeloid-biased HSCs, potentially contributing to immune alterations observed with aging. These studies also demonstrate that changes in functional potential of HSCs during development and aging correlate with changes in gene expression and growth factor requirements. Additionally, HSCs demonstrate epigenetic changes over time. Similar detailed studies in humans still need to be performed, but the murine studies raise the prospect that there will be similar changes in the functional potential of human HSPCs and that this may in part be the cause of the relative lymphoid deficiency in the aging population. In humans, the discovery that mutations in genes encoding epigenetic regulators are relatively common with aging, and affect the same genes recurrently mutated in myeloid malignancies, further demonstrates the importance of epigenetic control of normal hematopoiesis.

Somatic mutation(s) in hematopoietic stem/progenitor cells can drive clonal expansion (also known as clonal hematopoiesis) in the absence of clinical or laboratory evidence of hematologic malignancy has been recently described in great detail and occurs at relatively high frequency in an aging population (~10% to ~30%) and in patients who received cytotoxic chemotherapy for lymphoma (~30%). The most common genes involved in the aging population include *ASXL1, DNMT3A*, and *TET2*, with the additional enrichment of *PPM1D* and *TP53* genes in the cytotoxic chemotherapy–exposed population. The somatic mutations generally lead to a loss of function of 1 allele. A high variant allele frequency—a marker of the clonal hematopoietic cell burden in the blood—correlates with outcomes including increased deaths related to cardiovascular disease (strokes and heart attacks) and increased incidence of hematologic malignancies or therapy-related myeloid neoplasms. The biological connection between mutant hematopoietic cells and cardiovascular disease is believed to be in part related to unchecked inflammation, which has been studied in *TET2*-deficient mouse models. Interestingly, in a nonhematologic cancer patient population, therapy-related clonal hematopoiesis was associated with increased deaths due to progression of the primary nonhematologic cancer. In addition to unchecked inflammation playing an important role in tumor microenvironment and tumor

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

progression, the role of impaired immunogenic cancer surveillance and cell death should be further explored.

Hematopoiesis: a hierarchical differentiation cascade

Hematopoietic stem and progenitor cells represent a highly heterogeneous population defined by 2 key properties: variable ability to self-renew and variable ability to differentiate and generate mature blood cells. At the top of the hierarchy are HSCs. They have extensive self-renewal capability and differentiate into all blood cell types (Figure 1-6A). During normal steady-state hematopoiesis, adult HSCs cycle slowly and are relatively resistant to cytokine stimulation. Within the HSC compartment, long-term stem cells are usually quiescent and rarely divide. Short-term stem cells are more proliferative. The remarkable ability of HSCs, at the single-cell level, to reconstitute and maintain a functional hematopoietic system over extended periods of time in vivo demonstrates these key properties. Self-renewal allows HSCs to be transplanted between individuals, and the surviving HSCs engraft, proliferate, and differentiate for the life of the recipient. HSCs can be serially transplanted for many generations between recipients. The most primitive HSCs are rare, representing approximately 1 in 10^4 to 10^6 BM cells based on experimental models.

When HSCs are recruited into active hematopoiesis, they exit the G_0 phase of the cell cycle and undergo mitosis, leading to daughter cells with either identical cell fate (symmetric cell division) or different cell fate (asymmetric cell division). In the case of symmetric cell division, the two daughters can both retain HSC functions, or both display an activated differentiation program. In the case of asymmetric cell division, one daughter cell is a replicate of the parent cell (self-renewal) while the other daughter cell displays an activated differentiation program. This distinctive, asymmetric division process is the basis for long-term preservation of HSCs while enabling continued production of mature cells. The daughter cells that undergo differentiation proceed through a series of maturational cell divisions, culminating in the generation of progenitor cells.

Progenitor cells are also hierarchically arranged. As they differentiate from stem cells and through the progenitor hierarchy, they progressively lose self-renewal and become restricted in their differentiation potential such that more multipotent progenitors give rise to oligopotent and finally monopotent progenitors. Progenitors are highly proliferative and very cytokine responsive. New studies of murine hematopoiesis suggest that in contrast to transplantation, the source of most of the cells produced daily by the blood system during normal steady-state blood production is maintained by the continuing expansion of thousands of hematopoietic progenitor cells. Single-cell transplant studies in mice have also revealed a bypass pathway that produces long-term repopulating myeloid progenitors. This pathway may be operative under stress, as progenitor populations most readily respond to stress conditions in order to up- and downmodulate production of specific blood cell types. Progenitors differentiate into lineage-restricted precursor cells and eventually mature effector cells of the hematopoietic system. These mature lineages include erythroid cells for oxygen transport, myeloid and lymphoid cells that provide immune defense, and megakaryocytes and platelets essential for hemostasis.

Finally, our current view of the hematopoietic cellular hierarchy is changing based on data generated using newer technologies such as more advanced lineage tracing and single-cell omics profiling. Some key conceptual changes include the following: (1) hematopoiesis is a continuous differentiation process, not a stepwise process with discrete functional and phenotypic markers; (2) lineage segregation occurs earlier in the hierarchy at the HSC level rather than at the progenitor level; (3) hematopoiesis is not lineage-balanced, and HSCs are not homogenous and can be affected by aging, stress, and injury. Understanding of the hierarchical relationships between populations, the plasticity of commitment, and the nature of the functional potential of populations will increase as we better purify HSPC populations using newer technologies.

Informed by the rapidly expanding array of single-cell molecular profiling and functional assays, HSC differentiation is increasingly portrayed as a differentiation landscape. Unlike the more conventional hierarchical differentiation tree, these landscapes are informed by the observation that changes in gene expression occur in a gradual or continuous fashion. It is unclear however, whether there are nevertheless sharp boundaries in cellular functionality, as an individual cell differentiates and traverses these molecular landscapes (Figure 1-6B).

Phenotypic characterization and isolation of HSPCs

Attempts to purify stem cell populations have used a combination of approaches based on physical and biologic properties and cell surface marker expression of HSPCs. Early work on murine BM revealed that the transplantable HSCs copurified with lymphocytes and led to the idea that HSCs are morphologically indistinguishable from lymphocytes. Density gradient separation, such as separation with a Ficoll or Percoll gradient, is commonly used as a preenrichment step in stem cell purification protocols. Progenitor cells cycle actively, whereas HSCs are

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

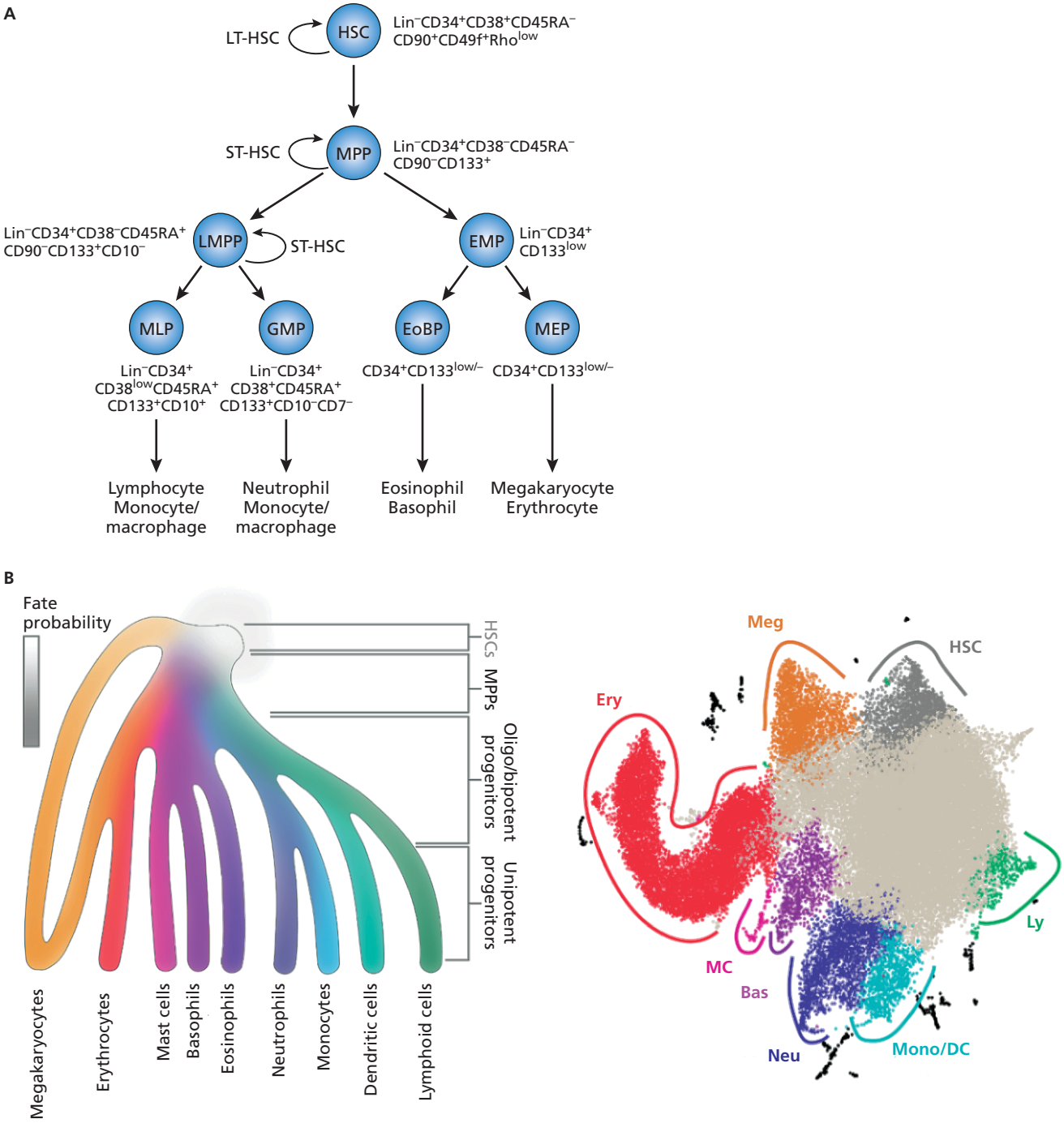


Figure 1-6 Alternative hierarchal maps of human hematopoiesis. (A) In the classical model of human hematopoiesis, a self-renewing HSC gives rise to a multipotent progenitor (MPP) that bifurcates into a common myeloid progenitor (CMP) and a common lymphoid progenitor (CLP), which then eventually give rise to mature myeloid and lymphoid cells respectively. According to more recent data, the MPP appears to give rise to a lymphoid primed multipotent progenitor (LMPP), both of which may have short-term self-renewal capacity, and an erythromyeloid progenitor (EMP), differentiated by CD133 expression. The hierarchy following the LMPP further bifurcates into a multilymphoid progenitor (MLP) and a granulocyte macrophage precursor (GMP). The EMP gives rise to a precursor that produces eosinophils and basophils (EoBP) and a megakaryocyte/erythroid progenitor (MEP). It is highly likely that as our ability to functionally and molecularly interrogate HSPCs at a single-cell level improves, our understanding of the hierarchy will be further refined. Surface antigen phenotypes of human hematopoietic stem and progenitor cells are indicated. The immunophenotypes of the EMP, EoBP, and MEP are not as well characterized as the other progenitors. LT-HSC, long-term HSC; ST-HSC, short-term HSC. (B) Schematic of a more gradual and flexible view of fate determination (left) inferred from single-cell transcriptomic landscapes of bone marrow hematopoiesis (right). Bas, basophils; Ery, erythroid precursors; Ly, lymphocytes; MC, mast cells; Meg, megakaryocytes, Mono/DC, monocytes/dentritic cells; Neu, neutrophils

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

relatively quiescent. This difference has been exploited in techniques for HSC enrichment in mouse and human systems. Treatment of mice with the antimetabolite agent fluorouracil markedly reduces progenitor cells, while relatively sparing populations enriched in HSC activity. More recently, considerable progress has been made prospectively isolating HSPCs using flow cytometry and cell surface markers. However, it is difficult to compare different immunophenotyping strategies with respect to quantifying the purity of the HSC population, as various factors such as source of HSCs (umbilical cord versus marrow), route of transplant (intravenous vs intrafemoral), the type of immunodeficient mouse used for xenografting, and pretransplant manipulation of the cells can all affect the outcome.

Figure 1-7 schematically illustrates how HSCs and progenitors are isolated and tested for function. Hematopoietic tissues are isolated then disassociated through mechanical or enzymatic disruption and then labeled with panels of fluorescently conjugated antibodies. The cell populations can then be analyzed and separated by fluorescence-activated cell sorting. Early studies to isolate human HSCs found that approximately 1% of human BM cells express CD34. Isolation of $CD34^+$ cells enriches for HSPCs, which have hematopoietic engraftment potential when transplanted into irradiated nonhuman primates. Similarly, human $CD34^+$-selected cells contain stem cells capable of fully reconstituting the lymphohematopoietic system in humans after myeloablative chemotherapy and radiation therapy. Although it is clear that the CD34-expressing population contains a long-term, repopulating HSC, there is some evidence of an upstream deeply quiescent $CD34^-$ HSC that gives rise to the $CD34^+$ HSC. Approximately 5% to 25% of $CD34^+$ cells also express low to moderate levels of CD90. CD90 expression by human hematopoietic cells decreases with differentiation, and most lineage-restricted progenitors are $CD34^+CD90^{+/low}$ cells. Additional studies demonstrate that human HSCs do not express mature cell lineage markers (Lin^-) or CD45RA or CD38. Isolation of $Lin^-CD34^+CD38^-CD45RA^-CD90^+$ cells provides a relatively easy method to sort for putative HSCs based on in vitro and in vivo studies. However, this remains a heterogeneous population. Sorting for the integrin CD49 further enriches for HSCs. Others have combined some of these markers with ability of HSCs to efflux dyes (eg, the mitochondrial dye rhodamine 123). HSCs, but not progenitor cells, express high levels of the verapamil-sensitive multidrug-resistance membrane efflux pump (P-glycoprotein), which confers resistance to multiple chemotherapeutic agents. This pump also excludes certain fluorescent dyes, such as rhodamine 123 or Hoechst 33342.

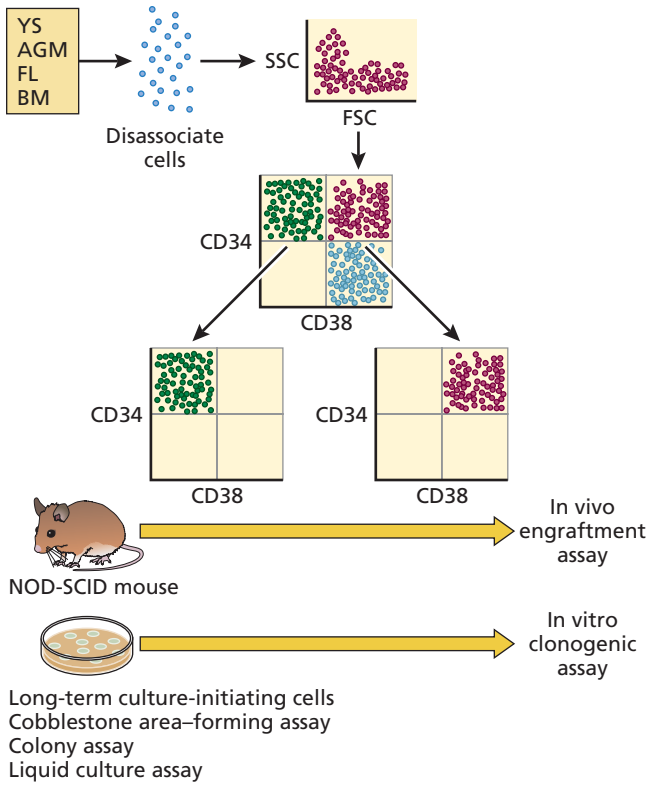


Figure 1-7 Isolation of HSCs. This figure shows how HSCs can be isolated from different sources. Cells are initially disassociated and stained with multiple antibodies. They are then analyzed and sorted with a fluorescence-activated cell sorter. Here mononuclear live cells are separated in gate 1. These live cells are then analyzed for CD34 and CD38 expression. Those live cells that are $CD34^+CD38^-$ are enriched for stem cell potential (green circles). Further purification can be undertaken on the basis of additional cell surface markers such as CD45RA, CD90, CD49f, and efflux of dyes (eg, rhodamine). To test the functionality of isolated (sorted) cells, cells can be tested in in vivo assays (transplanted into immunodeficient mice such as the NOD-SCID mouse model) and in vitro in long-term culture (long-term culture–initiating cell culture assay and cobblestone-area forming assay), clonogenic colony assays and liquid culture assays. FL, fetal liver; FSC, forward side scatter; SSC, side scatter. Adapted from Hoffbrand AV et al, Color Atlas of Clinical Hematology, 4th ed. (Mosby, Ltd; 2010), with permission. © John Wiley & Sons Ltd.

By using these dyes in combination with flow cytometry, it has been possible to identify a population of hematopoietic cells with low dye retention, called side population (SP) cells. Although this population is markedly enriched for HSCs, SP cells still represent a heterogeneous mix and are not equivalent to pure HSCs. Although the SP phenotype has been useful in characterizing HSCs (and possibly other non-HSCs) isolated from mice, this characteristic has not translated as easily into the human system.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

Downstream of the long-term repopulating HSCs, one study has shown that CD133 expression differentiates various progenitor populations, such that high CD133 defines the lymphoid-primed multipotent progenitor (LMPP) that matures into lymphoid, monocyte/macrophage, and neutrophil lines. In contrast, $CD133^{low/-}$ progenitors differentiate into megakaryocytes, erythrocytes, eosinophils, and basophils (Figure 1-6). In mice, the immunophenotype of c-Kit^+, Thy-1^+, Lin^- (a cocktail of surface markers found on mature cells of distinct lineages), and Sca-1^+ (called KTLS cells) enriches for cells with HSC activity. Flk2 expression can be used to distinguish long-term repopulating HSCs (LT-HSCs; $Flk2^-$) from short-term repopulating HSCs (ST-HSCs; $Flk2^+$). Other protocols have used the signal lymphocyte activation molecule (SLAM) family receptors CD150, CD244, and CD48 to isolate murine HSCs that are highly purified as $CD150^+CD244^-CD48^-$. The number of murine HSCs estimated using either the KTLS/$FLK2^-$ or the SLAM immunophenotype is ~10,000 HSCs/mouse. Of note, the SLAM phenotype does not translate for isolation of human HSCs.

Though current HSPC populations are still impure, they have allowed isolation of cell populations of defined functionality and are useful to identify genes and signaling pathways that mediate human HSPC differentiation. Isolation of distinct HSPC populations is also beginning to permit careful dissection of the hierarchical relationships between different blood cell populations. This is essential in describing the cellular basis of normal hematopoiesis. In turn, this is critical when trying to understand (1) the normal cellular compartments where genetic and epigenetic changes initially occur in hematological diseases (ie, the disease-initiating cell populations); (2) the cell compartments where subsequent mutations/epigenetic change is acquired during disease evolution and how this changes the hematopoietic hierarchy; and (3) the cell populations that propagate hematopoietic disease. Advances in cell sorting, genetic analysis, and other technologies are making analysis of HSCs increasingly precise. There is also a coming together of the more conventional flow cytometry and newer single cell molecular profiling methods, principally through the development of new single-cell multiomic methods, which permit the simultaneous analysis of cell surface marker proteins as well as the cell's transcriptome, all at single-cell resolution. Such research efforts are vital to anchor new profiling data sets in the wealth of existing experimental data on the functionality of conventionally defined hematopoietic stem and progenitor cells.

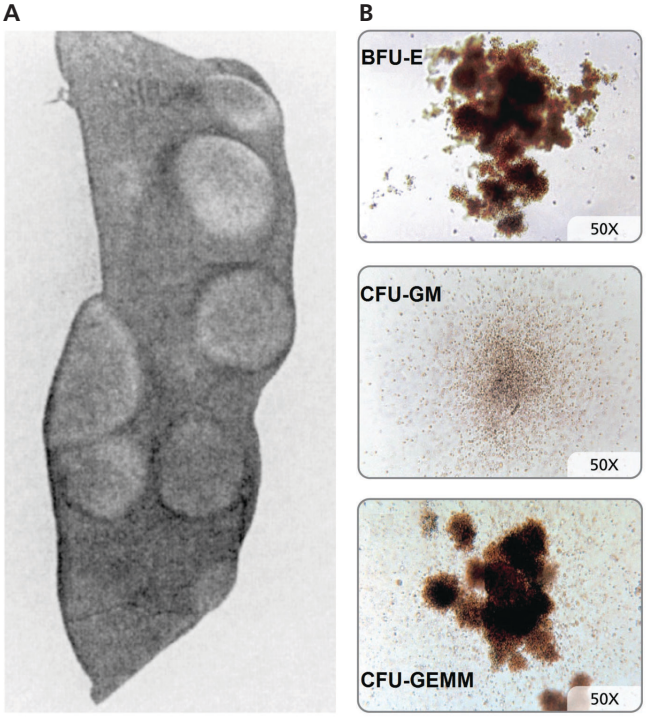
Stem/progenitor cell assays

A number of in vitro and in vivo assays have been developed to test HSPC function. It is useful to have background knowledge of them, as the assays have defined HSPC populations.

Colony-forming assays

The identification of a cell capable of in vivo clonal differentiation by Till and McCulloch in 1961—for example, in spleen colonies (Figure 1-8A)—prompted other groups to develop a simple quantitative assay for the growth and differentiation of single-cell suspensions of mouse BM in vitro. When hematopoietic cells were cultured in a semisolid medium (typically, soft agar or methylcellulose), discrete colonies were formed and included cells in multiple stages of differentiation (Figure 1-8B). In line with the properties observed for CFU-S, it was subsequently established that colonies generated in vitro could be initiated

Figure 1-8 (A) Spleen colony-forming unit (CFU-S) assay. Macroscopic splenic hematopoietic colonies arising from the CFU-S stem/progenitor cell 14 days after injection of murine BM into lethally irradiated mice. Reproduced from Williams DA, Stem cell model of hematopoiesis. In: Hoffman R et al, eds. Hematology: Basic Principles and Practice (Churchill Livingstone; 1995), with permission. (B) Examples of colony-forming assays of human hematopoietic progenitor cells. These include burst-forming unit erythroid (BFU-E), CFU granulocyte/macrophage (CFU-GM), and CFU granulocyte/erythroid/macrophage/megakaryocyte (CFU-GEMM). Reprinted from STEMCELL Technologies (www.stemcell.com), with permission.



Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

by the proliferation of a single colony-forming cell (CFC). In contrast to the self-renewal potential of most CFU-S, colonies grown in vitro displayed more limited ability to proliferate in secondary cultures. Therefore, CFCs were suggested to define a population of committed progenitors. That is why today we define progenitors as those cells that can form a colony in an in vitro colony-forming assay.

Long-term bone marrow culture

Attempts to develop procedures that mimic the marrow microenvironment resulted in the development of long-term BM cultures. In these assays, formation of an adherent stromal cell layer, which produces and deposits an extracellular matrix meshwork, is a prerequisite for the development and maintenance of hematopoietic cells. In association with the feeder layer, hematopoietic cells proliferate and differentiate over several months in culture-releasing clonogenic and mature cells. The ongoing production of these cells is the result of differentiation and proliferation of primitive cells. In recognition of their method of detection, these cells have been called long-term culture-initiating cells (LTC-ICs). They represent primitive immature hematopoietic cells that can be assayed in vitro. The presence of LTC-ICs can be detected by assaying for the presence of CFUs in cultures maintained for a minimum of 5 weeks. Beyond this point, any CFCs (progenitor cells with shorter survival time) initially present in the culture should have disappeared through differentiation or death, and those detected are the result of differentiation by LTC-IC. LTC-ICs are not necessarily true HSCs, and limits of these assays make it difficult to know whether these cells are capable of definitive long-term reconstitution and maintenance of hematopoiesis in vivo. In vivo studies can be expensive and cumbersome, however, and they come with their own caveats. Therefore, LTC-IC studies provide a reasonable in vitro surrogate assay for immature human hematopoietic cells with similar functions to HSCs.

Transplantation assays

The definitive assay for mouse HSC activity is the ability to provide long-term (>4 months) repopulation of all blood lineages of myeloablated host mice. Human HSCs cannot be similarly identified, however, except in a clinical study. Therefore, xenograft models have commonly been used as another surrogate assay for human HSCs. This work originally involved transplanting human hematopoietic cells into severe combined immune-deficient (SCID) mice. However, more immunodeficient and/or radioresistant mouse strains such as NOD-SCID/IL-2R$\gamma^{-/-}$ (commonly termed NSG or NOG mice) or NOD/Rag1$^{-/-}$/IL-2R$\gamma^{-/-}$ (NRG) or further engineered to express human cytokine or other molecules that aid cell survival (for example, IL-3, Steel factor, GM-CSF termed NSGS mice; human thrombopoietin or human SIRP1a) are now used for these analyses. Human cells giving long-term engraftment in these immunodeficient models are considered more primitive and clearly distinct from prior multipotent primitive human hematopoietic cell populations identified using in vitro methodology. Other transplantation models have been developed; for example, using fetal sheep for xenografts or nonhuman primates for autologous transplantation studies (often using gene transfer into putative HSCs). Additionally, zebrafish (where HSCs are located in the kidney and not BM) have become a well-utilized model system of hematopoiesis. Zebrafish are amenable to medium- and high-throughput analyses and have been used to identify many genes and soluble factors that regulate hematopoiesis.

Single-cell transplantation studies can identify clonal mouse and human hematopoietic cells with the ability to mediate long-term, multilineage engraftment. Several recent research approaches, however, have demonstrated that HSCs are a heterogeneous cell population. Earlier studies used phenotypic cell surface antigens to distinguish HSCs with long-term engraftment ability (LT-HSCs) and those that mediate just short-term engraftment (ST-HSCs). More recent studies (primarily in mice), however, have demonstrated that some HSCs have an inherent lineage bias that is at least partly maintained through serial transplantation in the mouse.

Pluripotent stem cells and hematopoiesis

Mouse embryonic stem cells (mESCs) were first isolated in 1981. mESCs have been proven invaluable for studies of basic mammalian developmental biology, including hematopoietic development. Unlike adult stem cells (such as HSCs), ESCs are able to undergo self-renewal indefinitely in culture, yet maintain the ability to form all somatic cell lineages (including hematopoietic cells). Studies with mESCs have been invaluable to identify genes that regulate hematopoietic development through gene deletion and/or manipulation. Additionally, diverse hematopoietic cell populations can be derived from mESCs in vitro and allow for interrogation of specific genetic and cell signaling pathways that regulate development of specific hematopoietic cell lineages. Notably, attempts to derive HSCs capable of long-term multilineage engraftment have largely failed using mESCs that have not been manipulated genetically. However, overexpression of certain TFs, most notably HoxB4, has been shown

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

to produce hematopoietic cells capable of engraftment in syngeneic recipients. Other TFs have had similar effects.

hESCs were first described in 1998. Like mESCs, hESCs can be maintained indefinitely as a self-renewing population in culture, yet maintain the ability to form all somatic cell populations. hESCs also have been used to investigate human hematopoiesis. Indeed, key areas of human hematopoiesis are distinct from the murine system. For example, human globin genes undergo 2 switching events during embryonic-fetal development, whereas the mouse undergoes only 1 switching event. hESCs also have raised considerable interest because of the potential for using these cells to produce large amounts of human cells and tissues suitable for research purposes, stem cell transplantation, or transfusion medicine. For example, there has been considerable interest in using hESCs to produce red blood cells or platelets as an adjunct to the living donor blood supply. Additionally, the potential to produce HSCs from hESCs is of great interest. To date, however, although most mature blood cell populations have been produced from hESCs, it has not been possible to demonstrate long-term engraftment of HSCs by transplantation into immunodeficient mice. Genetic manipulation and overexpression of TFs (such as HoxB4) effective in the murine system has not been similarly effective in the human system. Considerable efforts to identify strategies to improve development of HSCs from hESCs are ongoing.

iPSCs are another important cell population (see also the previous section "Disease-in-a-dish"). Briefly, iPSCs can be derived from various somatic cell populations, typically by expression of a limited number of "reprogramming genes," such as *OCT4*, *SOX2*, *KLF4*, and *c-MYC* in human cells, capable of converting somatic cells into cells that look and function like embryonic stem cells. These studies were first done in mouse cells in 2006 and subsequently in human cells in 2007. Like their ESC counterparts, iPSCs have been used to derive diverse hematopoietic cell lineages. Again, to date, HSCs with long-term engraftment potential have not been derived from iPSCs. This field will continue to mature, and there is considerable interest in deriving iPSCs from individuals with different genetic deficiencies to use this system as a human model of genetic disease. Using iPSCs, gene correction strategies, or other means to overcome the genetic defect can be evaluated. This may lead to effective therapies based on using iPSCs as a screening resource and would not require direct transplantation of iPSC-derived cells. The generation of disease stage-specific iPSCs in myeloid malignancies has provided insights into the cellular events demarcating the initiation and progression of transformation and a new platform for testing genetic and pharmacological interventions. Future developments may allow for derivation of iPSCs from individuals with hematologic or other diseases and use of these cells to produce autologous replacement cell populations. However, some key challenges that prevent broad translational impact of iPSCs as therapeutic sources in the human hematopoetic system remain, including unknown short-term and long-term functional capacity of HSCs derived from iPSCs, scalability of small molecule reprogramming methods to generate sufficient cells for therapeutic purposes, and high burden of safety and regulatory requirements and costs for clinical research studies and eventual clinical application involving iPSCs. Continued progress is expected in iPSC research, which may address some of these remaining barriers.

The hematopoietic niche

Hematopoietic cell development from HSPCs is regulated by signals provided by the BM microenvironment. The specific constituents of the microenvironment that influence blood cell development are being elucidated, but they can be categorized broadly as heterologous cells, such as mesenchymal cells; endothelial and neural cells; hematopoietic cells; and extracellular matrix. Mesenchymal cells include adipocytes, osteoblasts, leptin receptor–expressing ($LepR^+$) and nestin-positive cells. Some are part of the continuum of cells that produce bone, and some are perivascular without a clear role in skeletal biology. Both bone-related and perivascular mesenchymal cells have been shown to influence hematopoiesis.

The posttransplant BM niche is directly relevant to HSPC recovery following transplantation. In vivo imaging studies of the posttransplant niche demonstrated that some transplanted HSPCs are found close to the endosteal surface and osteolineage cells. HSPC-osteolineage cell colocalization in the posttransplant BM niche may be indicative of a regulatory relationship and single-cell RNA-Seq shows that HSPC-proximal osteolineage cells have a distinct RNA-Seq profile and can regulate HSPC quiescence. Nestin-positive mesenchymal stem cells are important for HSC persistence, and adipocytes variably regulate HSC numbers (Figure 1-9). Previously implicated as negative regulators of HSC number, recent work suggests adipocytes in long bones promote hematopoietic recovery after irradiation by being an important source of stem cell factor (SCF) (Figure 1-9).

The heterogeneous mesenchymal stromal cell (MSC) population plays a significant role in the hematopoietic niche. In human studies, cord blood cocultured with MSCs underwent a median 30-fold expansion of $CD34^+$ cells and resulted in significantly improved engraftment.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

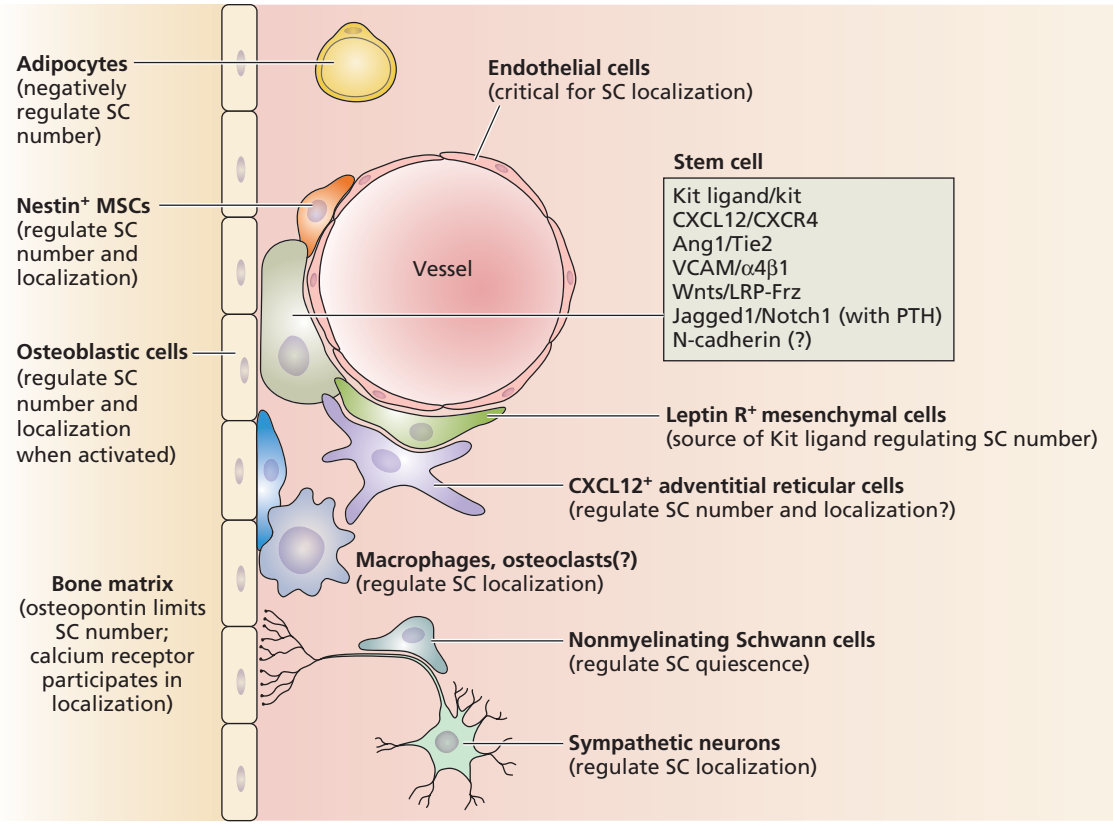


Figure 1-9 Bone marrow niche. HSCs localize to perivascular spaces, some of which are near the endosteal surface. A number of mouse models have been used to define specific cell types and gene products that, when manipulated, result in a change in HSC location or number. The cell types are indicated in the figure with the HSC function that their activity appears to modulate. The molecules involved are collected in the gray box, but which cells express these molecules is still being investigated. Some of the molecules are well defined (eg, kit ligand and CXCL12) while others are less well defined, and some (N-cadherin) are quite controversial.

In mice, LepR$^+$ cells represent the majority of MSCs. LepR$^+$ cells appear to be the main source of new osteoblasts and adipocytes in adult BM and form bony ossicles supportive of hematopoiesis in vivo. LepR$^+$ MSCs are the major source of the cytokines SCF that promotes proliferation of cells expressing the SCF-receptor c-kit, and chemokine (CXC motif) ligand 12 (CXCL12), which mediates adhesion of the HSCs in the BM niche. Conditional deletion of the *SCF* gene in LepR$^+$ cells leads to depletion of quiescent HSCs and conditional deletion of the gene encoding CXCL12 (*CXCL12,* also called *SDF1*) in LepR$^+$ cells leads to HSC mobilization.

Other cell types, such as neural cells of the sympathetic nervous system and nonmyelinated Schwann cells, also play a role in HSC support or localization. The sympathetic nervous system mediates circadian modulation in the number of HSCs moving from BM to bloodstream on a daily basis. Mature hematopoietic cells are also thought to influence HSC function in the BM. Specifically, macrophages help regulate HSC mobilization into blood and T cells are thought to influence HSC engraftment and provide relative protection from immune attack. Megakaryocytes have been shown to be important for maintaining HSC quiescence. Secreted factors, including CXCL4, TGFb1, and thrombopoietin, have also been implicated in this role. Therefore, a complex admixture of cells participates in what is designated as the stem cell niche.

The niche serves several functions important for hematopoiesis. The first is the regulation of stem cell self-renewal, a process that requires expression of molecules, such as SCF and members of the WNT family. The second is control of the number of stem cells, a parameter that is regulated in part by specific extracellular matrix proteins, such as osteopontin, a negative regulator of HSC number. The third is the coordinated regulation of proliferation and differentiation of HSCs, a process that some mouse models have indicated can go awry with changes in the niche and cause myeloproliferative or myelodysplastic phenotypes. The fourth is cell localization, a process that is important in the context of harvesting stem cells

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

by mobilization into the blood or delivery of transplanted HSCs to enable engraftment.

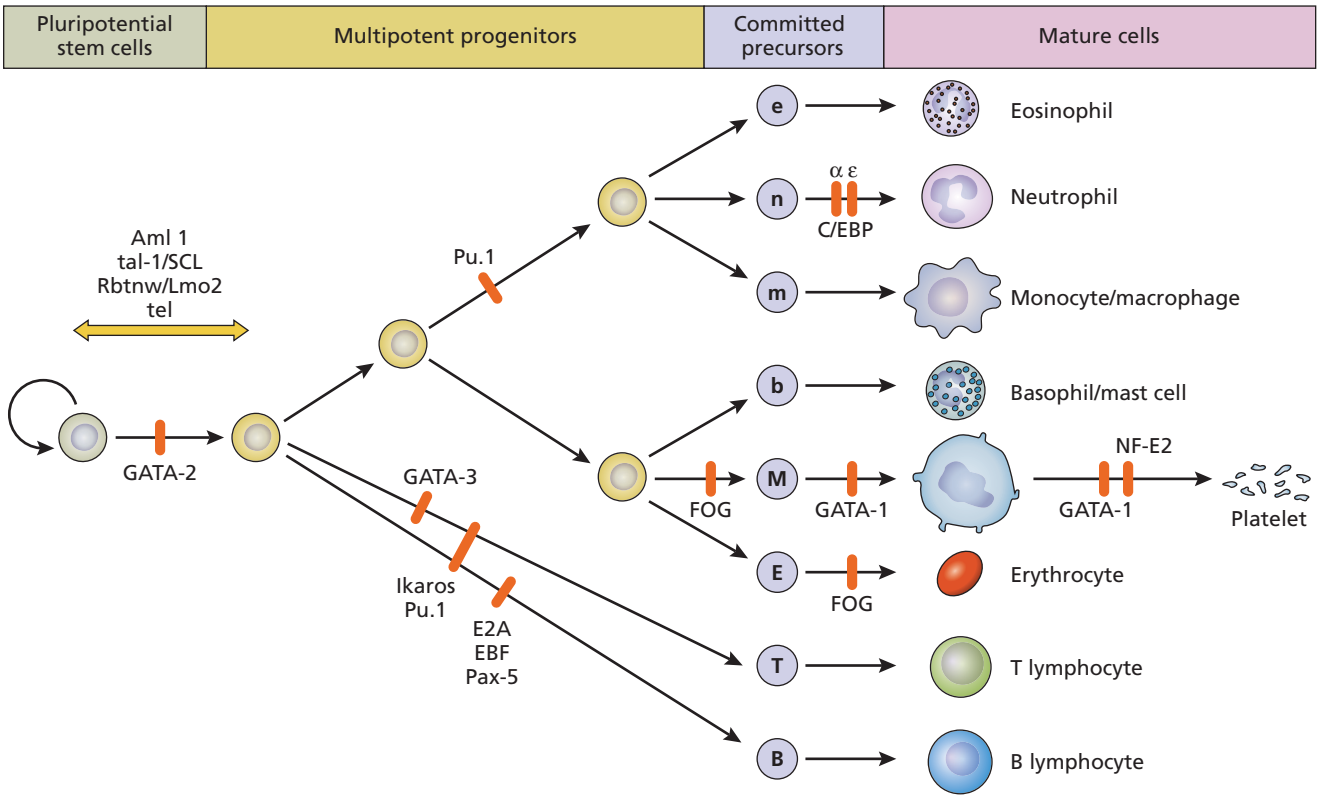
Thus, the HSC niche is a critical aspect of the regulated production of blood cells throughout life. It is a complex tissue in which multiple cell types and extracellular matrix proteins contribute to balance the molecular cues that govern HSC number, self-renewal, and differentiation. By unraveling how stem cells enter and leave the niche, methods to mobilize stem cells for clinical harvest have been defined. Ongoing efforts to improve stem cell function and engraftment in the niche and to discern how the niche contributes to disease are contexts in which manipulation of the niche may provide therapeutic potential.

Regulation of hematopoietic differentiation

A complex network of transcription factor and growth factor signaling pathways regulates HSPC self-renewal, lineage commitment, and differentiation. Among TFs, those that are expressed exclusively in blood cells or have restricted tissue-specific patterns of expression play important roles in regulating blood production. Furthermore, acquired driver mutations of these TFs are often pathogenic in hematological malignancies such as lymphoma and leukemia. The importance of these TFs is also underscored by the conserved role they play in hematopoiesis through evolution. Over the last 2 decades, this attribute has allowed the function of these TFs to be extensively investigated in animal models. In these models, genes encoding critical TFs have been deleted, modified, overexpressed, and misexpressed. The point of action of some of these TFs is shown in Figure 1-10. A thorough description of the function of these TFs is beyond the scope of this chapter. Some of the key points that arise from these studies follow:

1. TFs are divided into families that have similar protein domains.
2. TFs often have protein domains that bind DNA and protein domains that interact with other proteins (other TFs or proteins that control transcription).

Figure 1-10 A schematic representation of hematopoiesis and where key hematopoietic-specific transcription factors have nonredundant functions, as revealed by gene deletion studies in mice. Thus, for example, the transcription factors GATA2, AML1/RUNX1, TAL-1/SCL, LMO2/RBTN2, and TEL are all critically required in HSCs and loss of function of these genes causes a block (as indicated by the red bar) in hematopoietic differentiation at the HSC level. Similarly, deletions of the other transcription factors cause blocks later in hematopoiesis (as indicated by the orange bars).



Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

3. TFs work in combination with other TFs to activate and/or repress the expression of a large number of genes.
4. TFs are required at discrete stages of hematopoiesis. Any one TF can function at multiple stages within a single lineage and can function in more than one lineage.
5. Ultimately, TFs work in complicated networks that can be modeled much like semiconductor/computing networks. TFs work in negative feedback loops, feed-forward loops, and cross-antagonistic loops, to mention just 3 such types of interactions.
6. TFs regulate the cell's potential to make blood cells of different lineages and to proliferate, undergo apoptosis, and self-renew.

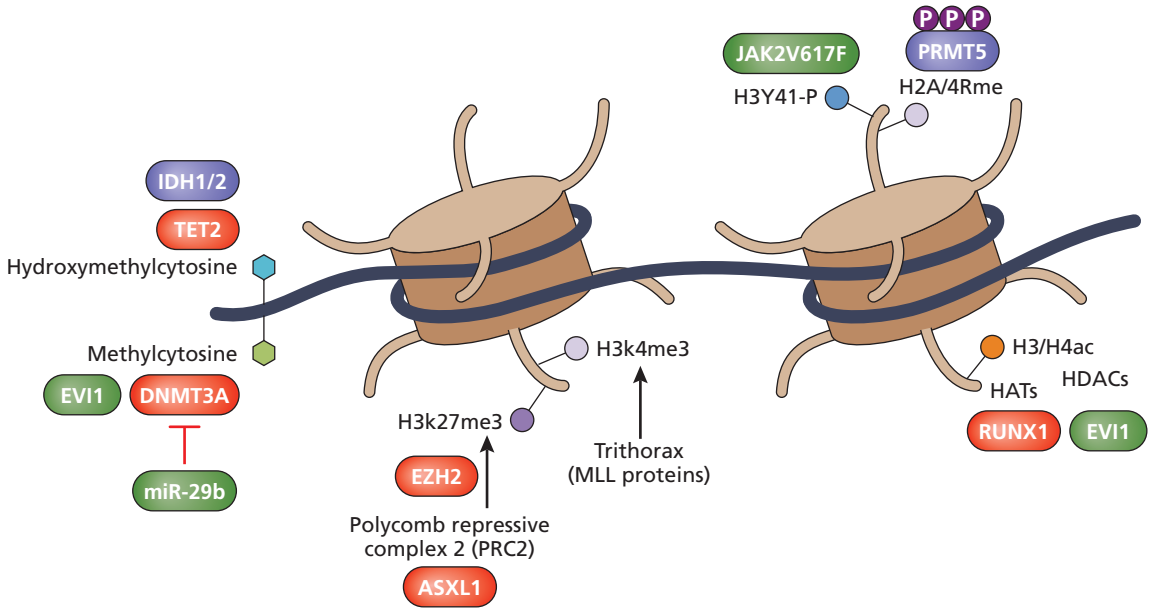
More specifically, the TFs SCL/TAL1 and LMO2 are required to specify HSCs from mesoderm. The TFs RUNX1 (AML1), TEL1, MLL, and GATA2 are required to maintain stem cells once they have been specified. In myelopoiesis, the TFs Pu.1, the C/EBP family (C/EBPa and C/EBPe) GFI-1, EGR-1, and NAB2 all promote the granulocyte-macrophage lineage programs. GATA2 is required in stem/early progenitor cells but is also required for mast cell differentiation and in the early phases of megakaryocyte-erythroid lineage maturation. Working with GATA2 to promote erythropoiesis and megakaryopoiesis are GATA1, FOG1, SCL, EKLF, p45NF-E2, and FlI-1. In early lymphopoiesis, the TF Ikaros is required. In B lymphopoiesis, the TFs E2A (and its family members), EBF, and PAX5 are required; and finally, the TF BLIMP1 is necessary for plasma cell formation. In T-cell maturation, Notch signaling activates the TF CSL, which works with the TFs GATA3, T-BET, NFATc, and FOXP3. Some of the many TFs known to be implicated in the pathogenesis of human leukemia include the TFs SCL/TAL1, MLL, RUNX1, LMO2, PU.1, C/EBPa, PAX5, E2A, and GATA1.

In addition to TFs, proteins that modulate the epigenetic profile of cells (eg, regulate DNA methylation and histone modifications) and regulate splicing of RNA are also commonly mutated either through loss-of-function or gain-of-function mutations. This also suggests that these proteins play critical roles in normal HSPC differentiation. Examples of proteins that act as epigenetic modulators are provided in Figure 1-11.

Summary

Hematopoiesis involves a tightly regulated set of developmental stages, from HSCs to hematopoietic progenitor cells to mature blood cells, which provide all the key functions of the hematopoietic system. Hematopoietic reconstitution during BMT is mediated by a succession of cells at various stages of development. Immediately

Figure 1-11 Epigenetic modifiers that are altered in sequence or expression in hematopoietic malignancies. Genes outlined in red represent loss-of-function mutants, whereas those outlined in green represent gain-of-function or overexpressed genes. The genes outlined in blue represent mutants that have acquired novel (neomorphic) function. These epigenetic modifiers may impact on DNA methylation (DNMT3A, TET2, IDH1, IDH2), histone (H3 or H4) methylation, or histone acetylation, thereby modifying gene expression across a wide range of targets. ac, acetyl; HAT, histone acetyltransferase; HDAC, histone deacetylase; me, methyl.



Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

following transplantation, more mature cells contribute to repopulation. With time, cells at progressively earlier stages of development contribute, with the final, long-term repopulation provided by long-lived multipotent HSCs. Research in induced pluripotent stem cells and in the BM niche offers the potential for a greater understanding of disease biology and for novel therapies to emerge. Clonal hematopoiesis is an age- and chemotherapy-associated disease syndrome associated with clonal expansion of mutant hematopoietic cells and adverse health outcomes. The pathologic mechanisms of how mutations in epigenetic regulatory genes and in other genes disrupt normal hematopoiesis, immune system, and inflammatory pathways in humans remain to be determined.

KEY POINTS

Development

- Hematopoiesis develops in distinct waves during development.
- Definitive HSCs first develop within the embryo in specialized regions of the dorsal aorta and umbilical arteries, then seed the fetal liver and BM.
- HSC characteristics differ based on their site of development and the age of the organism.

Key features of HSCs

- Ability, at the single-cell level, to reconstitute and maintain a functional hematopoietic system over extended periods of time in vivo.
- Self-renewal capacity for life of organism or after transplantation.
- Multipotency: the ability to make multiple types of blood cells.
- Relative quiescence: the ability to serve as a deep reserve of cells to replenish short-lived, rapidly proliferating progenitors.
- In vivo transplantation models are currently the only reliable assays of HSC activity and function.
- Acquired somatic mutations in hematopoietic stem cells can lead to clonal expansion (clonal hematopoiesis), a phenomenon associated with adverse health outcomes including cardiovascular disease and hematologic malignancies.

Key features of hematopoietic progenitor cells

- Inability to maintain long-term hematopoiesis in vivo because of limited or absent self-renewal.
- More rapid proliferation and cytokine responsiveness, enabling increased blood cell production under conditions of stress.
- Display lineage commitment, and thereby limited cell-type production.

Key features of the HSC niche

- Anatomically and functionally defined regulatory environment for HSCs.
- Modulates self-renewal, differentiation, and proliferative activity of HSCs, thereby regulating stem cell number.
- Niche function is important in maintaining HSC integrity; therefore, niche dysfunction may contribute to hematopoietic disease.
- Niches for HSCs are dynamic, changing during development and with physiologic stress.
- HSCs naturally traffic into and out of the niche, a feature that can be exploited for stem cell transplantation or harvesting, respectively.

Glossary

alleles Alternative forms of a particular gene.

allele-specific oligonucleotide An oligonucleotide whose sequence matches that of a specific polymorphic allele. For example, oligonucleotides matching the sequence of unique immunoglobulin or T-cell receptor gene rearrangements that are used for polymerase chain reaction detection of minimal residual disease.

alternative splicing Selective inclusion or exclusion of certain exons in mature RNA by utilization of a varied combination of splicing signals.

antisense oligonucleotides Oligonucleotides with a base sequence complementary to a stretch of DNA or RNA coding sequence.

ATAC-seq High-throughput sequencing approach to measure DNA accessibility.

capping Addition of the nucleotide 7-methylguanosine to the 5′ end of mRNA. This is a structure that appears to stabilize the mRNA.

ChIP-Seq A combination of chromatin immunoprecipitation followed by next-gen sequencing used to identify protein-DNA interactions.

chromatin A complex of genomic DNA with histone and nonhistone proteins.

chromosome A large linear DNA structure tightly complexed to nuclear proteins.

***cis*-acting regulatory elements** Sequences within a gene locus, but not within coding sequences, that are involved in regulating the expression of the gene by interaction with nuclear proteins.

clonal Arising from the expansion of a single cell.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

coding sequence The portion of the gene contained within exons that encodes the amino acid sequence of the protein product.

codon The 3-nucleotide code that denotes a specific amino acid.

comparative genomic hybridization (CGH) A technique allowing for the detection of subtle chromosomal changes (deletions, amplifications, or inversions that are too small to be detected by conventional cytogenetics techniques).

complementary Sequence of the second strand of DNA that is determined by strict purine–pyrimidine base pairing (A-T; G-C).

complementary DNA (cDNA) Double-stranded DNA product from an RNA species. The first strand is synthesized by reverse transcriptase to make a DNA strand complementary to the mRNA. The second strand is synthesized by DNA polymerase to complement the first strand.

constitutive promoter A promoter that drives high-level expression in all tissues.

copy number variant A segment of DNA at least 1 kb in length that varies in copy number between individuals.

CRISPR/Cas (clustered regularly interspaced short palindromic repeat/Cas-based RNA-guided DNA endonucleases) A technology that combines the Cas DNA nuclease with the sequence-specific DNA recognition module of CRISPR to create targeted genetic alterations in DNA.

cytogenetics The study of the chromosomal makeup of a cell.

Dicer A component of the processing mechanism that generates microRNAs and siRNAs.

Drosha A component of the processing mechanism for formation of microRNAs.

enhancer A *cis*-acting regulatory sequence within a gene locus that interacts with nuclear protein in such a way as to increase the expression of the gene.

enzyme-linked immunosorbent assay (ELISA) A method used to detect and quantify proteins (such as peptides, proteins, antibodies and hormones) using an antibody linked to an enzyme.

epigenetics Changes in gene expression caused by mechanisms other than alteration of the underlying DNA sequence. Includes DNA methylation and histone modification. The changes are heritable in daughter cells but can be modified pharmacologically (eg, methyltransferase inhibitors, histone deacetylase inhibitors) or by normal enzymatic processes.

exome The set of all protein-coding portions of genes (exons) in the genome.

exon The portion of a structural gene that encodes protein.

flanking regions DNA sequences lying 5′ and 3′ of a structural gene that frequently contain important regulatory elements.

flow cytometry A method to quantify protein expression on cells using antibodies, where the cells with bound antibodies are passed in a single suspension through a machine with a laser to detect abundance.

fluorescence in situ hybridization (FISH) High-resolution mapping of genes by hybridization of chromosome spreads to biotin-labeled DNA probes and detection by fluorescent-tagged avidin.

frameshift mutation A mutation within the coding sequence of a gene that results from deletion or insertion of a nucleotide that disrupts the 3-base codon structure of the gene, thereby altering the predicted amino acid sequence of the protein encoded by that gene.

gene A functional genetic unit responsible for the production of a given protein, including the elements that control the timing and the level of its expression.

gene expression profile Analysis of the global expression of a collection of cells using hybridization of mRNA to microarrays.

gene regulation A process controlling the timing and level of expression of a gene.

genetic code The system by which DNA encodes specific proteins through 3-nucleotide codons, each encoding a specific amino acid.

genomics The study of the entire DNA sequence of organisms and interactions among various genetic loci.

Hi-C A next-generation sequencing approach to identify long-range chromatin interactions genome-wide.

homologous recombination Alteration of genetic material by alignment of closely related sequences. In targeting genes by homologous recombination, plasmids that contain altered genes flanked by long stretches of DNA that match the endogenous gene are introduced into embryonic stem cells. A rare recombination event will cause the endogenous gene to be replaced by the mutated gene in the targeting plasmid. This is the means by which knockout mice are obtained.

immunohistochemistry (IHC) A method of detecting proteins in tissue sections using antibodies linked to substrates that allow for visual detection of antibody-protein binding abundance in situ.

imprinting A genetic process in which certain genes are expressed in a parent-of-origin–specific manner.

induced pluripotent stem cells (iPSCs) A type of pluripotent stem cell derived from a somatic cell that is generated by exposing the somatic cell to factors that reprogram it to a pluripotent state.

intron An intervening sequence of noncoding DNA that interrupts coding sequence contained in exons.

knockin mouse A mouse in which nucleotides have been inserted into the mouse genome to allow expression of protein not normally encoded by the mouse genome.

knockout mouse A mouse in which both of the copies of a gene have been disrupted by a targeted mutation. Such mutations are achieved by homologous recombination using plasmids

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrillaki on 10 December 2022

containing the mutated gene flanked by long stretches of the normal endogenous gene sequence. Mice that are heterozygous in the germline for the targeted allele can be bred to generate mice that lack both copies of the normal (wild-type) gene.

leucine zipper Leucine-rich side chains shared by a group of transcription factors that allow protein-protein and protein-DNA interactions.

linkage mapping Analysis of a gene locus by study of inheritance pattern of markers of nearby (linked) loci.

methylation DNA modification by addition of methyl groups to cytosine residues within genomic DNA. Hypermethylation of clustered CpG groups in promoter regions (CpG islands) is a characteristic of transcriptionally inactive DNA; reduction in methylation is generally associated with increased transcriptional activity.

microarray A glass slide or silicon chip on which cDNAs or oligonucleotides have been spotted to allow for the simultaneous analysis of expression of hundreds to thousands of individual mRNAs. Hybridization of labeled cDNAs from a tissue of interest allows the generation of a gene expression profile.

microRNAs (miRs) Small RNA molecules encoded in the genomes of plants and animals. These highly conserved, approximately 21-mer RNAs regulate the expression of genes both by changing stability of mRNAs as well as by translational interference.

missense mutation A mutation within the coding sequence of a gene that results from a single nucleotide change that alters the encoded amino acid, leading to a change in protein function.

next-generation (next-gen) sequencing Massively parallel sequence production from single-molecule DNA templates.

noncoding sequences DNA sequences that do not directly encode protein.

nonsense mutation A nucleotide change converting an amino acid coding codon to a stop codon.

nonsense-mediated decay Nonsense mutation (premature stop codon) of one allele of an mRNA may result in degradation of the abnormal mRNA.

nucleotide A basic building block of nucleic acids, composed of a sugar moiety linked to a phosphate group and a purine or pyrimidine base.

oligonucleotide A short single-stranded DNA species, usually composed of 15 to 20 nucleotides.

oncogene Cellular gene involved with normal cellular growth and development, the altered expression of which has been implicated in the pathogenesis of the malignant phenotype.

partial uniparental disomy A situation in which two copies of a chromosome, or part of a chromosome, are derived from one parent and no copies derive from the other parent. In a somatic cell, this can result in progeny with two copies of the wild-type allele or two copies of the mutant allele.

polyadenylation Alteration of the 3′ end of mRNA by the addition of a string of adenosine nucleotides (“poly-A tail”) that appear to protect the mRNA from premature degradation.

polymorphism A phenotypically silent mutation in DNA that is transmitted from parent to offspring.

pre-messenger RNA Unprocessed primary RNA transcript from DNA, including all introns.

promoter Region in the 5′ flanking region of a gene that is necessary for its expression; includes the binding site for RNA polymerase II.

proteomics The systematic study of the entire complement of proteins derived from a cell population.

Proto-oncogens Normal functioning genes that when mutated can become oncogenes and thus, transform cells.

quantitative PCR PCR in which the product is quantitated in comparison to the PCR product resulting from a known quantity of template. This allows quantitation of the template in the reaction; it can, for example, allow an estimate of the degree of contamination with tumor cells in a cell population.

real-time PCR An automated technique for performing quantitative PCR using a fluorogenic reporter to detect levels of target sequences during early cycles of the PCR reaction.

reverse transcriptase An enzyme encoded by retroviruses that mediates conversion of RNA to complementary DNA.

reverse transcriptase polymerase chain reaction (RT-PCR) Amplification of RNA sequences by conversion to cDNA by reverse transcriptase, followed by the polymerase chain reaction.

ribosome A ribonuclear protein complex that binds to mRNA and mediates its translation into protein by reading the genetic code.

RNA expression array An array-based technique used to determine the abundance of each of the known mRNAs (the gene expression profile) in a group of cells.

RNA-induced silencing complex (RISC) A multiprotein complex that combines with microRNAs to target complementary mRNA for degradation or translation inhibition.

RNA polymerase II An enzyme that mediates transcription of most structural genes.

RNA-Seq Next-gen sequencing using RNA templates.

silencer A *cis*-acting regulatory sequence within a gene locus that interacts with nuclear protein in such a way as to decrease the expression of the gene.

single-nucleotide polymorphism (SNP) Naturally occurring inherited genetic variation between individuals at the level of single nucleotides.

small interfering RNAs (siRNAs) Small RNAs that act in concert with large multiprotein RISCs to cause cleavage of complementary mRNA or prevent its translation.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

Southern blotting Analysis of DNA by gel electrophoresis, transfer to nitrocellulose or nylon filter, and hybridization to single-stranded probe.

splicing The process by which intron sequences are removed from pre-mRNAs.

telomeres Nucleoprotein structures at the ends of chromosomes that protect chromosome ends from degradation and fusion.

termination codon One of three codons that signal the termination of translation.

***trans*-acting factor** A protein that interacts with *cis*-acting regulatory region within a gene locus to regulate transcription of that gene. Also called transcription factor.

transcription The process by which pre-mRNA is formed from the DNA template.

transcription activator–like effector nucleases (TALENs) Artificial restriction enzymes with sequence-specific DNA-binding activity that can be utilized to create specific genetic alterations in DNA.

transcription factor A protein that interacts with *cis*-acting regulatory region within a gene locus to regulate transcription of that gene. Also called *trans*-acting factor.

transfer RNA (tRNA) Small RNA molecules that bind to the ribosome and covalently bind specific amino acids, allowing translation of the genetic code into protein.

transgenic mouse A mouse that expresses an exogenous gene (transgene) introduced randomly into its genome. Linearized DNA is injected into the pronucleus of a fertilized oocyte, and the zygote is reimplanted. Resultant mice will carry the transgene in all cells.

translation The process by which protein is synthesized from an mRNA template.

translocation breakpoint Site of junction of two aberrantly juxtaposed (translocated) chromosomal fragments.

tumor suppressor A gene that promotes tumor development when deleted or inactivated.

zinc finger A structural feature shared by a group of transcription factors. Zinc fingers are composed of a zinc atom associated with cysteine and histidine residues; the fingers appear to interact directly with DNA to affect transcription.

zinc finger nucleases Artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain for use in creating specific genomic alterations in DNA.

Bibliography

Molecular techniques in hematology

George LA, Sullivan SK, Giermasz A, et al. Hemophilia B gene therapy with a high-specific-activity factor IX variant. *N Engl J Med*. 2017;377(23):2215-2227. *Successful example of the use of gene therapy for the treatment of hemophilia.*

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-351. *Review of next-generation sequencing techniques and methodologies.*

Hammond SM. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett*. 2005;579(26):5822-5829. *A review describing the discovery of the RNA interference pathway and discussion of future lines of work.*

Hanna J, Wernig M, Markoulaki S, et al. Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science*. 2007;318(5858):1920-1923. *Using a humanized sickle cell anemia mouse model, these investigators showed that mice can be rescued after transplantation with corrected hematopoietic progenitors obtained in vitro from autologous iPS cells.*

Hughes T, Branford S. Molecular monitoring of BCR-ABL as a guide to clinical management in chronic myeloid leukaemia. *Blood Rev*. 2006;20(1):29-41. *A description of the use of real-time PCR to monitor BCR-ABL transcripts in chronic myeloid leukemia.*

Jackson HJ, Rafiq S, Brentjens RJ. Driving CAR T-cells forward. *Nat Rev Clin Oncol*. 2016;13(6):370-383. *A review of CAR T cells.*

Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet*. 2014;15(10):647-661. *A review of DNA cytosine methylation and other modifications of DNA cytosines.*

Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32(4):347-355. *A review of CRISPR-Cas technology for genome-editing.*

Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. 2017;14(9):865-868.

Volpi EV, Bridger JM. FISH glossary: an overview of the fluorescence in situ hybridization technique. *BioTechniques*. 2008;45(4):385-409. *A succinct review of the many FISH strategies available.*

Hematopoiesis

Becker AJ, McCulloch EA, Till JE. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature*. 1963;197(4866):452-454.

Bonig H, Papayannopoulou T. Mobilization of hematopoietic stem/progenitor cells: general principles and molecular mechanisms. *Methods Mol Biol*. 2012;904:1-14.

Méndez-Ferrer S, Lucas D, Battista M, Frenette PS. Hematopoietic stem cell release is governed by circadian oscillations. *Nature*. 2008;452(7186):442-447.

Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm of stem cell biology. *Cell*. 2008;132(4):631-644.

Park D, Sykes DB, Scadden DT. The hematopoietic stem cell niche. *Front Biosci*. 2012;17(1):30-39.

Purton LE, Scadden DT. Limiting factors in murine hematopoietic stem cell assays. *Cell Stem Cell*. 2007;1(3):263-270.

Silberstein L, Goncalves KA, Kharchenko PV, et al. Proximity-based differential single-cell analysis of the niche to identify stem/progenitor cell regulators. *Cell Stem Cell*. 2016;19(4):530-543.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022

Sutherland HJ, Lansdorp PM, Henkelman DH, Eaves AC, Eaves CJ. Functional characterization of individual human hematopoietic stem cells cultured at limiting dilution on supportive marrow stromal layers. *Proc Natl Acad Sci*. 1990;87(9):3584-3588.

Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res*. 1961;14(2):213-222.

Zhou BO, Yu H, Yue R, et al. Bone marrow adipocytes promote the regeneration of stem cells and hematopoiesis by secreting SCF. *Nat Cell Biol*. 2017;19(8):891-903.

Clonal hematopoiesis

Coombs CC, Zehir A, Devlin SM, et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. *Cell Stem Cell*. 2017;21(3):374-382.e4.

Cull AH, Snetsinger B, Buckstein R, Wells RA, Rauh MJ. Tet2 restrains inflammatory gene expression in macrophages. *Exp Hematol*. 2017;55:56-70.e13.

Fuster JJ, MacLauchlan S, Zuriaga MA, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science*. 2017;355(6327):842-847.

Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477-2487.

Gibson CJ, Lindsley RC, Tchekmedyian V, et al. Clonal hematopoiesis associated with adverse outcomes after autologous stem-cell transplantation for lymphoma. *J Clin Oncol*. 2017;35(14):1598-1605.

Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.

Jaiswal S, Natarajan P, Silver AJ, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N Engl J Med*. 2017;377(2):111-121.

Peng J, Yang Q, Li AF, et al. Tet methylcytosine dioxygenase 2 inhibits atherosclerosis via upregulation of autophagy in ApoE−/− mice. *Oncotarget*. 2016;7(47):76423-76436.

Zhang Q, Zhao K, Shen Q, et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature*. 2015;525(7569):389-393.

Ontogeny of hematopoiesis

Bertrand JY, Chi NC, Santoso B, Teng S, Stainier DYR, Traver D. Hematopoietic stem cells derive directly from aortic endothelium during development. *Nature*. 2010;464(7285):108-111.

Chen MJ, Yokomizo T, Zeigler BM, Dzierzak E, Speck NA. Runx1 is required for the endothelial to hematopoietic cell transition but not thereafter. *Nature*. 2009;457(7231):887-891.

Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat Immunol*. 2008;9(2):129-136.

Höfer T, Rodewald H-R. Output without input: the lifelong productivity of hematopoietic stem cells. *Curr Opin Cell Biol*. 2016;43:69-77.

Kumaravelu P, Hook L, Morrison AM, et al. Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver. *Development*. 2002;129:4891-4899.

North TE, Zon LI. Modeling human hematopoietic and cardiovascular diseases in zebrafish. *Dev Dyn*. 2003;228(3):568-583.

Stem cell characteristics

Baum CM, Weissman IL, Tsukamoto AS, Buckle AM, Peault B. Isolation of a candidate human hematopoietic stem-cell population. *Proc Natl Acad Sci USA*. 1992;89(7):2804-2808.

Copley MR, Beer PA, Eaves CJ. Hematopoietic stem cell heterogeneity takes center stage. *Cell Stem Cell*. 2012;10(6):690-697.

Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med*. 1996;183(4):1797-1806.

Kiel MJ, Yilmaz ÖH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*. 2005;121(7):1109-1121.

Kotini AG, Chang CJ, Chow A, et al. Stage-specific human induced pluripotent stem cells map the progression of myeloid transformation to transplantable leukemia. *Cell Stem Cell*. 2017;20(3):315-328.e7.

Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature 2018;553(7689):418-426.

Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science*. 2011;333(6039):218-221.

Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. *Science*. 1988;241(4861):58-62.

Wagers AJ, Sherwood RI, Christensen JL, Weissman IL. Little evidence for developmental plasticity of adult hematopoietic stem cells. *Science*. 2002;297(5590):2256-2259.

Downloaded from http://ashpublications.org/books/book/chapter-pdf/1908571/chapter_1.pdf by Eleni Gavrilaki on 10 December 2022